L1 transfer effects in L2 grammatical gender processing of late bilinguals

DISSERTATION

zur Erlangung des akademischen Grades

Doctor philosophiae

(Dr. phil.)

Eingereicht an der
Philosophischen Fakultät II
der Humboldt-Universität zu Berlin

von **Anna Renner (M. Sc.)**

Datum der Verteidigung: 14.4.2014

Präsident der Humboldt-Universität zu Berlin:
Prof. Dr. Jan-Hendrik Olbertz

Dekanin der Philosophischen Fakultät II:
Prof. Dr. Helga Schwalm

Gutachterinnen:
1. Prof. Dr. Katharina Spalek (Humboldt-Universität zu Berlin)
2. Prof. Dr. Rosemarie Tracy (Universität Mannheim)

Contents

Abstract

The present dissertation investigates first language (L1) transfer effects in second language (L2) grammatical gender processing of late bilinguals. The term "late bilinguals" refers to people who learn an L2 after childhood, in contrast to "early bilinguals". Research has shown that L2 acquisition after childhood is usually less successful than during childhood. The reasons for this are still debated. Interestingly, some aspects of a language are more affected by age than others and sometimes negative age effects can be overcome at very high proficiency levels. One of the structures especially affected by AoA is grammatical gender. One possible explanation for L2 processing difficulties in late bilinguals is negative transfer from the L1. For this reason, the present thesis focuses on L1 transfer effects in L2 gender processing of late bilinguals. Transfer arises because all languages of a speaker are activated and compete for selection at all times. One aim of this thesis is to describe which factors or combination of factors influence L1 gender transfer. Regarding L2 gender processing in general, different factors have been shown to affect performance. Among these are language proficiency of the subjects, task demands, and syntactic distance of the agreeing elements. Gender transfer has been shown to be affected by factors such as characteristics of the L1 gender system, transparency of the L2 gender system, and form similarities of nouns in L1 and L2 (cognates vs. noncognates). Research further suggests that transfer might be influenced by similarity of the languages and symmetry of the gender systems. Besides this, little is known on how gender transfer is mediated by L2 proficiency and by the complexity of the L2 gender system. In the present thesis, a behavioral and an ERP experiment were conducted. Gender transfer was investigated across different language pairs with gender systems of varying complexity and transparency. Experimental tasks differed in task demands and syntactic structures with varying agreement distances were used. Language proficiency of subjects was also manipulated. Based on my findings, I was able to identify which factors and which combination of factors increase or decrease gender transfer and to describe gender transfer as the result of a complex interplay of a combination of various factors.

Acknowledgments/Danksagung

Ich danke meiner Doktormutter Prof. Dr. Katharina Spalek für die Möglichkeit diese Dissertation zu schreiben, für ihr großes Interesse an meinem Thema und die zuverlässige und umfangreiche Betreuung in allen fachlichen Fragen.

Ich danke meiner Familie, Freunden und Kollegen für ihre Unterstützung, Ermutigung, Ratschläge und für das Korrekturlesen.

Ich danke außerdem der FAZIT-Stiftung, dem DAAD, dem Gleichstellungsfonds der Humboldt-Universität zu Berlin und dem Gleichstellungsfonds des Institutes für Deutsche Sprache und Linguistik für die finanzielle Förderung meiner Arbeit.

List of abbreviations

AoA	Age of acquisition
CP	Critical period
CPH	Critical period hypothesis
DP Model	Declarative/procedural model
ERPs	Event-related potentials
EEG	Electroencephalogram
fMRI	Functional magnetic resonance imaging
GJT	Grammatical judgment task
LDT	Lexical decision task
L1	First language
L2	Second Language
NP	Noun phrase
PDW	Projekt Deutscher Wortschatz
PET	Positron emission tomography
PNT	Picture naming task
RTs	Reaction times
SD	Standard deviation

0. Introduction

Bilingualism and even multilingualism are normal phenomena in numerous countries around the world. In many parts of the world, especially in Africa and Asia, several languages are spoken within a country and children grow up learning more than one language in a natural manner (Bhatia & Ritchie, 2006). Consider for example Nigeria, a country with about 250 ethnicities (Konrad Adenauer Stiftung, 2012), each with its own language, and four official languages: English, Hausa, Igbo, and Yoruba (Nigeria Embassy Berlin, 2012). In most Western countries, however, the picture is different. In everyday life only one language is used by the majority. Therefore, learning another language is a task usually begun later in life and through formal instruction. As a consequence, becoming fluent in another language is not seen as something natural but as something special and is considered a challenging endeavor. This is understandable as research confirms that learning a language later in life is usually less successful than in childhood (Johnson & Newport, 1989).

Nevertheless, due to increasing globalization, job demands, and migration, also in the Western world more and more people are forced to learn a second language (L2), oftentimes as adults. Especially the European Union (EU) as an economic and political union of 27 countries with 23 official languages encourages its citizens of all ages to learn foreign languages. This is supported by the promotion of language learning in school, vocational-educational trainings as well as mobility programs for all age groups (*Barcelona European Council*, 2002). The need for this becomes apparent when one considers that currently already 10 million Europeans work in other EU countries and the number is growing (*Multilingualism: An Asset for Europe and a Shared Commitment (COM) 2008 566 final*, 2008). In addition, due to increasing immigration from outside the EU also people from many different non-EU countries now live in the EU and are faced with the challenge of learning another language. The fact that in 2007 for the first time a single commissioner with multilingualism as their only responsibility was appointed illustrates the increasing importance the European Commission attributes to the promotion of multilingualism in the EU (*Key Data on Teaching Languages at School in Europe*, 2008).

Unfortunately, there seems to be a great discrepancy between the present day practical need of many people, in the EU and other parts of the world, to learn a language at a later age and the scientific finding that language learning after childhood is usually less successful. In order to deepen our understanding of the nature of the difficulties of late L2 learning more research is needed. It is already a well-known fact that some aspects of a language are more severely affected by age effects than others. One of the domains where mastery greatly deteriorates with higher age of acquisition[1] (AoA) is grammar[2], especially morphosyntax. Grammatical gender as a part of morphosyntax, for example, is relatively difficult to learn and can remain prone to errors even after many years of L2

[1] Some researchers make a distinction between "acquisition" and "learning", using the former when the learner is exposed to a language in a natural environment as opposed to formal instruction, in which case the latter term is used. As type of language exposure is not essential for the present thesis, both terms will be used interchangeably.

[2] Note that even though in linguistics the term *grammar* includes at least the fields of morphology, syntax, and phonology, in many (neuroscientific) articles, the term *grammar* is used to refer only to morphology and (morpho)syntax. In the present thesis, I will explicitly distinguish between morphosyntax and syntax when necessary, while in other places I will use the term *grammar* to refer to morphology and (morpho)syntax together.

exposure. Especially the German gender system poses many problems in L2 acquisition. Determiner agreement errors, such as *der Tür* (*the-$_{masc}$ door-$_{fem}$*) instead of the correct form *die Tür* (*the-$_{fem}$ door-$_{fem}$*), to name just one example, are common among L2 learners of German. In many cases, these errors do not greatly impair understanding. Nevertheless, gender errors hamper communication because they are a source of distraction. In addition, next to foreign accent, speech errors have been shown to make people come over as less professional and can even lead to discrimination (Eisenstein, 2008; Hosoda, Nguyen, & Stone-Romero, 2012; Hosoda & Stone-Romero, 2010). Since grammatical gender is so difficult to master for L2 learners, it is important to find out more about the acquisition of grammatical gender and how this develops in the course of L2 learning. Investigating the acquisition of L2 gender as an example of an especially difficult structure may also yield knowledge that can have a wider use for improving language learning and teaching strategies. For these reasons, the present thesis deals with the learning of grammatical gender.

So what is grammatical gender? According to the much-cited definition by Hockett , "genders are classes of nouns reflected in the behavior of associated words" (Hockett, 1958, p. 231; as cited in Corbett, 1991).

For instance, in the example given previously, the fact that the German noun *Tür* is feminine is reflected in it taking the associated feminine article *die*. And as Corbett further expands "There are various other ways in which nouns could be grouped: those denoting animals, those which are derived from verbs, those whose stem has three syllables or more, those whose stress changes from singular to plural. These groupings are not genders […] because they do not determine other forms beyond the noun; they are classifications internal to the class of nouns." (p. 4). This means that, almost by definition, gender can be difficult to identify as an intrinsic characteristic of a noun. It can only be recognized when words like determiners, adjectives, or verbs take different forms in association with the different noun classes.

Furthermore, another characteristic of gender is that there is a lot of variation across languages (Corbett, 1991). Some languages have a very complex gender system, whereas others, like the Turkish language (Comrie, 1999, p. 458), have no gender system at all. In some languages gender plays a central role, while it has little importance in others. Also, the rules by which gender is assigned differ widely: Some languages have a strict semantic system, others a predominantly or partly semantic system. Yet other languages have a formal, that is, morphological or phonological system[3] which can vary in transparency and ambiguity. Additionally, a combination of semantic, morphological, and phonological rules is possible (as in the case of German, e.g.). Furthermore, if a language has gender, four genders are common and up to twenty are possible (Corbett, 1991, p. 5). Common categories are masculine, feminine, neutral, animate and inanimate. Hence, considering the apparent lack of consistency and the therefore ostensible arbitrariness among languages when it comes to gender, it appears somewhat less surprising when gender poses an obstacle on the way to L2 proficiency. This seems especially true in the case of complex, intransparent, or ambiguous systems. Nevertheless, the difference between "the ease with which native speakers assign nouns to genders" and "the difficulty experienced by foreign learners of many gender languages" (Corbett, 1991, p. 8) is remarkable.

[3] But, as Corbett (1991, p. 63) points out, even if it is sometimes difficult to recognize, gender always has a basis in semantics. Dixon (1968), on the other hand, states that, as in many Bantu languages there is no semantic correlation at all with sex, the term noun class would be more suitable. Gender could be seen as a special case when there are two or three classes which correlate with sex.

One reason for difficulties with L2 gender and, among others, also a possible reason for detrimental age effects in L2 learning are (negative) transfer effects from the first language (L1) which hinder proper L2 acquisition and processing. The present thesis investigates L2 grammatical gender processing by looking at transfer effects from the L1 and how these change throughout the learning process. Concerning L2 grammatical gender processing and L1 gender transfer, many questions can be asked: Under which circumstances does L1 gender transfer occur? Which factors enhance gender transfer? How does L1 gender transfer change in the course of language learning? Does transfer decrease with increasing proficiency? How is L1 gender transfer modified by the transparency and complexity of the L2 gender system? Is L1 transfer possible into an L2 with a very simple gender system? With the present thesis investigating L1 gender transfer effects in "late" bilinguals, that is, bilinguals who learned their L2 after childhood, I will try to contribute to the answer of these questions.

To this end, I report the results of two experiments investigating L1 gender transfer in different language pairs and proficiency groups, with different tasks and methodologies:

Experiment 1 (chapter 4) investigates gender transfer in both directions between a Romance and a Germanic language, namely, Spanish and German. Both languages have grammatical gender but the gender systems differ, especially regarding their transparency and complexity. Online comprehension and production is investigated in noun phrases (NPs) and bare nouns, as well as offline gender assignment. Subjects were native speakers of German and Spanish who learned their respective L2s[4] (Spanish and German) after childhood and were thus "late bilinguals", as well as native control groups. Proficiency effects are also investigated.

Experiment 2 (chapter 5) investigates if gender transfer is possible from German, a gendered language, into English, an ungendered language. This time, gender transfer in anaphor resolution in sentence comprehension is investigated using event-related potentials (ERPs). Subjects are German–English[5] late bilinguals. Proficiency effects are also investigated.

Hence, there are three topics that can be identified as central to the present thesis: AoA and proficiency effects in L2 learning and the consequences for late bilinguals, grammatical gender in L2 learning, and L1 transfer effects in L2 learning. Before turning to the experiments and their results, in the next three chapters I will provide the reader with theoretical background on these three topics:

Chapter 1, "Age of acquisition effects in L1 and L2 learning" deals with AoA effects in L1 and L2 learning and how different domains (phonology, (morpho)syntax, and the lexicon) of a language are differently affected, pointing out that morphosyntax, and as such, grammatical gender is one of the areas that is most severely impaired. The role of L2 proficiency and two models explaining AoA effects and other phenomena are also discussed.

[4] It is possible that for some subjects, Spanish and German were their third or fourth languages. As in the present thesis and in most studies discussed in the present thesis it is not relevant to differentiate between L2, L3, or L4, the term L2 is used synonymous with "foreign language".

[5] In the present thesis I will use the following convention when talking about bilinguals: the language named first is the native language (in this case, German) and the language named second is the later acquired language (in this case, English).

Chapter 2, “The difficulty of grammatical gender”, discusses the problems associated with grammatical gender in L2 learning and the factors that mediate these difficulties.

Chapter 3, “L1 transfer effects in L2 lexicon, (morpho)syntax, and grammatical gender”, reviews the literature on interference and transfer studies in different language domains, explaining the current state of the art and the remaining questions. The last section in this chapter focuses on research on grammatical gender transfer which is where the present thesis comes into play.

After that, the experiments and their results will be summarized (chapters 4 and 5), followed by the final discussion (chapter 6).

1. Age of acquisition effects in L1 and L2 learning

As outlined in the introduction, the present thesis investigates L1 transfer effects in the acquisition of L2 grammatical gender in late bilinguals. In order to consider language learning in late bilinguals, it is necessary to first take into account what is special about late L2 acquisition and what the effects of age in language learning are. Therefore, the present chapter deals with the notion of a critical period (CP) and age of acquisition (AoA) effects in language learning. As we shall see, evidence regarding a CP in language learning is still controversial, while findings regarding AoA effects are more clear-cut. AoA effects in different language domains (phonology, (morpho)syntax, lexicon) as well as proficiency effects in L2 acquisition are discussed. At the end of the chapter, two models providing an explanation for AoA effects in language learning are presented.

1.1 The critical period hypothesis

First of all, what exactly is a CP and what does a CP for language acquisition imply? In his book "Psycholinguistics", Kess (1992) states that "The notion of a critical period is a familiar one in biology […]." And a CP entails that "Stimulation must take place during certain limited and prescribed critical periods of time, for if a particular behavior is not stimulated and responded to within a certain time frame, the behavior never fully or correctly emerges." (p. 268). Often cited examples of CPs in various areas of biology are song learning for sparrows (Marler & Peters, 1987; Marler, 1991), vision in cats (Hubel & Wiesel, 1959, 1962) and imprinting for chicken (Spalding, 1872) or geese (Lorenz, 1935). If, for example, song sparrows are not exposed to songs of their species at a critical age of 20 to 60 days after hatching (Marler, 1991) they will never learn how to sing properly. And geese adopt the first moving object they see within the first 24 to 36 hours of hatching as their mother and follow it around, no matter if it is really a goose or something else.

In a similar fashion, a CP has been postulated for language learning in humans. As we shall see, according to some researchers the CP of language acquisition is supposed to be responsible for impaired acquisition after childhood. The critical period hypothesis (CPH) for first language (L1) acquisition was first proposed by Penfield and Roberts (1959) and Lenneberg (1967). Penfield and Roberts state that "[…] a child´s brain has a specialized capacity for learning language – a capacity that decreases with the passage of years." Furthermore, "The brain of the child is plastic. The brain of the adult, however effective it may be in other directions, is usually inferior to that of the child as far as language is concerned." (p. 240). In the same vein, Lenneberg asserts that "Analogous to the question of how old must a child be before he can make use of the environment for language acquisition is the question of how young must an individual be before it is too late […]. There is evidence that the primary acquisition of language is predicated upon a certain developmental stage which is quickly outgrown at the age of puberty." (p. 142) and adds, "Thus we may speak of a critical period for language acquisition." (p. 179).
Lenneberg bases this claim, among others, on results on differential success rates in the recovery from aphasia in children and adults, completion of lateralization of the language function in the brain by puberty, studies on hemispherectomy at different ages, language development in retarded children, and the differential consequences of sudden deafness at different ages. Unfortunately, it remains somewhat unclear what his exact predictions concerning the success of language acquisition after the postulated CP are. This makes his claim difficult to falsify. It is uncertain whether he

considers L1 acquisition totally impossible or only impaired in certain aspects. These two possibilities will be referred to as a “strong version” and a “weak version” of the CPH, respectively.

Lenneberg did not attach much importance to conclusions drawn from the observations of so-called “wolf children” or “feral children” because he maintained that “The only safe conclusions to be drawn from the multitude of reports is that life in dark closets, wolves´ dens, forests, or sadistic parents´ backyards is not conducive to good health and normal development.” (p. 142). Nevertheless, the case of Genie (Curtiss, 1977; Fromkin, Krashen, Curtiss, Rigler, & Rigler, 1974) probably the most famous and well-studied case of a “feral child” provides evidence against at least a strong version of the CPH. Genie is a girl who suffered abuse and isolation from human contact since the age of 20 months till about 13 years of age when she was discovered. At that time, she was not able to speak. Genie was able to make up for some of her lacking language skills but did not attain the same proficiency as normally developed children in the time period of 7 years during which the research was conducted. Despite being very communicative she never completely mastered grammar rules, had problems with morphology and her semantic ability outperformed her syntactic ability. Fromkin et al. (1974) concluded that the study of Genie provided a challenge to the CPH. Even if her language attainment at the end of the research project was still different from normal adults, she nevertheless succeeded in mastering at least some English thereby showing that L1 acquisition is possible beyond the CP in any case at least up to a certain degree. So the case of Genie and other cases of feral children (e.g., Victor of Aveyron; Lebrun, 1980) plead for a weaker form of the CPH, that is, that L1 acquisition is also possible after the onset of puberty but possibly not as complete and successful as when speech develops earlier.

Evidence from feral children dates back to the 70s but there is also more recent evidence challenging a strong version of the CPH from, for example, a study with a deaf-born boy who started to learn verbal Spanish when he was fitted with hearing aids at the age of 15 (Grimshaw, Adelstein, Bryden, & MacKinnon, 1998). He learned how to speak but only with severe deficits in verbal comprehension and production, which the authors see as evidence in favor of the CP hypothesis[1]. Yet, according to a strong version of the CPH no language acquisition would have been possible at all. So it remains unclear how strong the supposed CP effects in L1 acquisition are. As investigation of the phenomenon is difficult and the underlying biological mechanisms potentially responsible for the CP are not yet completely clear, there even remain some doubts whether a CP applies to human language acquisition at all. As we shall see in the next section, CP effects in L2 acquisition are especially controversial.

1.2 L2 acquisition

Penfield and Roberts (1959) and Lenneberg (1967) extended their CPH based on neurological observations to L2 acquisition, assuming that the same maturational constraints influencing brain development and plasticity would also impinge on L2 acquisition. Penfield and Roberts state that “When new languages are taken up for the first time in the second decade of life, it is difficult, though not impossible, to achieve a good result. It is difficult because it is unphysiological.” (p. 255). Lenneberg asserts that “Most individuals of average intelligence are able to learn a second language after the beginning of their second decade, although the incidence of “language-learning-blocks “rapidly increases after puberty. […] This does not trouble our basic hypothesis on age limitations

[1] It has to be noted that whether evidence is interpreted as against or in favor of the CPH depends on how strong the CPH is interpreted by the author.

because we may assume that the cerebral organization for language learning as such has taken place during childhood […].” (p. 176). So for L2 acquisition it seems that both Penfield and Roberts, as well as Lenneberg formulated a weaker version of the CPH. Even if, as in the case of L1 acquisition, the expected learning outcomes of late L2 learners are not specified more precisely, it is possible to investigate the question of a CP for L2 acquisition mainly because larger sample sizes can be obtained. The question of whether there is a CP in L2 acquisition is usually approached by two types of paradigms (Ortega, 2009, pp. 18-19): a) correlational studies, typically involving a grammatical judgment task (GJT) and typically testing immigrants to the US, correlating AoA (usually equal to age of arrival) with grammatical judgment scores and b) comparisons between high-proficient late L2 speakers and early L2 speakers or native speakers on, for example, a GJT investigating whether these two groups behave differently and whether there are cases of native-like attainment in late learners.

Correlational studies

In the case of correlational studies, the usual finding is that L2 performance deteriorates with higher AoA. However, it is important to answer the question whether the correlations found are best explained by AoA or CP effects. AoA effects can be distinguished from CP effects by whether AoA effects continue to persist after puberty (Birdsong & Molis, 2001; Hakuta, Bialystok, & Wiley, 2003). As Hakuta et al. (2003) point out, if there really is a CP, the correlation between age and attainment should be different during and after the CP (p. 31). So if a gradual decline in proficiency with increasing AoAs is found across all age groups this would point to general age effects in language learning rather than CP effects (for a critique on this approach cf. Flege, Yeni-Komshian, & Liu, 1999, p. 80). It is also important to consider whether other sociolinguistic or educational factors influence performance because, as stated by Hakuta et al. (p. 32), if a supposedly purely biological mechanism, such as the CP is at work, those factors should not play a role. Furthermore, AoA has to be carefully separated from amount of exposure to the language, as these two factors are easily confounded.

The correlational research conducted into CP effects has yielded diverse and sometimes contradictory results. One of the first studies using the correlational approach and showing CP effects in L2 acquisition was the much-cited study by Johnson and Newport (1989). In a GJT, they tested the knowledge of different grammatical structures in Korean and Chinese immigrants. A significant decline in L2 proficiency was found for immigrants who had arrived to the US after the age of 16, compared to immigrants who had arrived before the presumed offset of the postulated CP. Test scores on the GJT correlated linearly with age of arrival until puberty, while after puberty test scores were very low, highly variable and the correlation with AoA disappeared. However, more recently, Birdsong and Molis (2001) replicated the study with the same material with Spanish-speaking immigrants but came to a different conclusion because L2 attainment continued to negatively correlate with AoA even after the end of the presumed CP. They also reported some cases of L2 learners reaching native-like proficiency. Furthermore, they found that success of L2 attainment was also mediated by factors such as L1–L2 similarities and L2 use.

In another study, Patkowski (1980) sought to control for possibly confounding sociolinguistic factors by ensuring that all subjects had been in an “optimal position” to acquire L2: Participants´ time of residence was 6 - 35 years and all were highly educated and studying or employed. Nevertheless, attained level of (syntactic) proficiency (as determined through the evaluation of oral interviews) was strongly correlated with age of arrival. Furthermore, it was shown that the development of proficiency before the chosen cut-off point of 15 years was different from the post-puberty development. In his conclusion, however, Patkowski favors the notion of a sensitive period over a CP which holds that native-like proficiency in *all* areas of a language is impossible to achieve for older learners, whereas very high levels of language proficiency in *some* areas do not necessarily conflict with this hypothesis (cf. also Oyama, 1976). But again, the results of the study by Patkowski are at odds with the results of Hakuta et al. (2003; also Wiley, Bialystok, & Hakuta, 2005) who in their study with data from 2.3 million immigrants to the US with Spanish or Chinese language backgrounds found no difference between the regression lines for L2 attainment on age of arrival on the two sides of the critical age point. There was thus no evidence for a CP but evidence for mere linear age effects. Furthermore, effects of level of education were found.

In a correlational study, Birdsong (1992) found that, as a group, English L1 speakers living in France differed from French native speakers on a variety of tasks in French (interpretation of ambiguous sentences, translation task, GJT). Because a native control group was tested as well, L2 speakers´ performance could also be compared to that of native speakers (the comparative approach explained above). It turned out that a majority of the L2 subjects actually performed within the native speaker range. These findings are at odds with Johnson and Newport (1989) and Patkowski (1980) where none, or only one of the L2 speakers, respectively, performed within the L1 speaker range. Furthermore, Birdsong´s data also allowed for examination of the correlation of age of arrival of the L2 subjects in France with performance measures and indeed, there was a negative correlation which was found to continue even after puberty. Thus, despite many studies finding evidence for AoA effects on L2 proficiency, with regard to the CP, results are inconclusive. In addition, the study by Birdsong even found evidence for late bilinguals performing like native speakers.

Comparisons between late L2 speakers and native speakers

Also the results of other studies comparing high-proficient late L2 learners to early L2 or native speakers are controversial. Some authors (e.g., Coppieters, 1987; Sorace, 1993) have found evidence of divergence and thus in favor of the CP in their studies on grammar intuitions of near-native speakers, while other authors did not (e.g., Montrul & Slabakova, 2003; White & Genesee, 1996). Coppieters, for example, interviewed native French and non-native subjects on their intuitions regarding the accuracy and meaning of sentences putting at test various grammatical contrasts. The author found that near-native speakers differed from native speakers in their acceptance and interpretation of sentences containing the two past tense forms imparfait vs. passé compose, the third person pronouns *il* and *ce* and pre-posed vs. post-posed adjectives. Sorace collected acceptability judgments of French and English L2 speakers of Italian as well as an Italian control group. Matter of investigation was “unaccusativity” in Italian, or put differently, selection of the correct auxiliary (*avere* vs. *essere*) in perfect tense. Also in this study, both non-native groups differed markedly from native speakers in their acceptability judgments.

White and Genesee (1996) selected a number of near-native L2 speakers of English after conducting interviews which were evaluated among others on pronunciation, grammatical errors, and fluency. The selected near-native speakers were compared to non-native speakers and an L1 control group on a GJT testing *wh*-movement and a question formation task. They found near-native speakers to be indistinguishable from native speakers, in terms of accuracy as well as speed. Nevertheless, the authors noted that “However, we do not intend to deny age effects altogether. […] in general, younger learners are more likely to achieve near-native proficiency than older learners.” (p. 258). Montrul and Slabakova (2003) tested the interpretation of tense aspect forms in L2 Spanish (preterite–imperfect contrast) which is difficult to grasp when the L1 does not distinguish between those two concepts. However, 70 % of the group of near-native speakers performed just like native speakers on a sentence judgment task.

All the same, as various researchers have argued, for a fair comparison of late L2 speakers with native speakers it would be preferable to compare late bilinguals to early bilinguals instead of monolinguals. That this is not taken into account can be seen as an important weakness of the non-correlational studies discussed. As stated by Grosjean (1989): “The Bilingual Is Not Two Monolinguals in One Person” (p. 3, for a discussion cf. also Ortega, 2009, pp. 26-27). Comparing the performance of late bilinguals to monolinguals bears the danger of confounding age effects in learning with effects that actually stem from bilingualism, that is, the fact that the languages of a bilingual inevitably interact (Birdsong, 2005, p. 320). As these cross-language effects are at the center of interest of the present thesis this topic will be discussed in more detail in chapter 3. Another shortcoming of the non-correlational studies is that, in contrast to the correlational studies, it is difficult to tell AoA from CP effects because the possibly differential relationship between age and attainment before and after the presumed end of the CP cannot be compared. As explained above, this approach provides the possibility to distinguish CP effects from AoA effects. Furthermore, one difficulty with the comparison of younger and older learners is the fact that AoA is oftentimes confounded by other variables, such as type and amount of L2 input or instruction. Clearly, quantity, quality, and content of L2 input are usually very different for children and adults.

As mentioned before, the obvious problems with the postulation of a CP for L2 learning has led some researchers (Oyama, 1976; Patkowski, 1980) to use the term CP only in the case of L1 acquisition and the more attenuated term “sensitive period” in the case of L2 acquisition. This choice of words expresses the notion that language learning limitations after puberty are probably stronger for L1 than for L2. A series of experiments testing the acquisition of American Sign Language (ASL) conducted by Mayberry suggests that this is indeed the case. Mayberry (1993) and Mayberry and Eichen (1991), for example, showed that AoA has a greater effect in L1 acquisition than in L2 acquisition. Mayberry and Eichen provided evidence that the long-term success and ultimate attainment of ASL as an L1 strongly depended on AoA, even after on average 42 years of exposure. The later subjects had acquired ASL, the worse they performed on a variety of linguistic tasks. Mayberry showed that subjects with English as an L1 who had learned ASL as an L2 after loss of hearing in late childhood performed significantly better than subjects who had learned ASL at the same age but as an L1. Other late learners of sign language who had normal hearing and normal acquisition of spoken English until they became deaf perform better than deaf-born signers who acquire ASL late. So it seems that AoA effects are stronger for L1 than for L2 acquisition. In a later study, Mayberry and Lock (2003) interpret these and other findings as support for the CP. Yet, it is unclear whether these effects can really be called CP effects according to the criteria mentioned by Hakuta, Bialystok, and Wiley (2003, discussed at the beginning of this section). But regardless of whether there is a CP or mere AoA effects in L2 learning, in both cases higher AoA appears to correlate with lower achievement of the L2.

In sum, whether there is a CP or sensitive period for language learning or not is still a matter of debate. Quite a few studies carefully investigating differences in correlation of AoA and proficiency before and after puberty, controlling for or manipulating other (sociolinguistic) factors of possible influence did not find evidence for a CP. However, AoA effects were found in most cases. So independent of whether these age effects meet the criteria for a CP (as discussed in section 1.2), and more important for the present thesis, in each case it seems clear that age has a detrimental effect on language learning. Furthermore, it is also clear that these age effects, while probably stronger for L1 than L2 acquisition, in some cases still greatly impair L2 learning. Hence, in order to find out more about the problems on the way to native-like attainment in L2 learning after childhood, the present thesis focuses on late bilinguals.

Although in the past quite a lot of research focused on the question of CP effects, present-day research is much more fine-grained and focuses more on the question of attainment in different language domains and similarities between early and late bilinguals. Research has become more domain-specific and as we shall see in the next section, it is evident that age effects are stronger for some areas of L2 learning than for others[2]. There are also neuroimaging techniques available now, allowing for more precise studies investigating processing as it unfolds online and identifying where or at what time point processing difficulties or differences to native speakers arise. Research on proficiency effects and changes in the course of language learning in late bilinguals have become another center of interest in the study of bilingualism. The present thesis builds on this development by investigating performance of late L2 learners of different proficiencies in just one very specific language domain, namely grammatical gender. Age effects in the learning of grammatical gender will be discussed in chapter 2. Yet, little is known on the reasons why grammatical gender is so difficult to master for late L2 learners. Research on L1 transfer of grammatical gender and how it develops in the course of learning can help to elucidate this matter, which is the approach taken in the present thesis. In the next section, I will discuss the different results for L2 learning in the various language domains in order to show that not all language domains are equally affected by AoA. This makes the question why grammatical gender is so difficult for late L2 learners all the more intriguing.

1.3 AoA effects in different language domains

Nowadays, there is wide agreement among researchers that different aspects of a language are differently affected by AoA (Birdsong, 2005b; Ortega, 2009). In the following, I will first review behavioral studies investigating AoA effects in phonology and (morpho)syntax. Then, I will discuss ERP studies investigating AoA effects in syntax, morphosyntax, and semantics, as well as ERP studies directly comparing (morpho)syntactic and semantic processing. The results of the ERP studies are important for the second experiment of the present thesis which also employed the ERP technique. Moreover, as grammatical gender is a part of morphosyntax, the results on AoA effects in (morpho)syntax are especially relevant for the present thesis.

1.3.1 Behavioral studies

Phonology

There is broad consensus on a CP for phonology. This CP is thought to end quite early, around age 6 (Flege et al., 1999). A famous example for the common fact that adult L2 learners retain a heavy accent in their L2 is the Polish author Joseph Conrad who wrote successful novels in his L2 English but preserved a heavy Polish accent throughout his life (Scovel, 1969). For this reason, this phenomenon became known as the “Conrad Phenomenon”. Genie´s case (Fromkin et al., 1974) is also a good example for the deterioration of speech abilities with higher AoA because she had special difficulties

[2] As Birdsong (2006) observed, this has led some researchers to propose “multiple critical periods” (p. 18) rather than just one general CP for language learning.

with the physical abilities needed for pronunciation. In a study of immigrants to the US Oyama (1976) found that age of arrival was the strongest predictor for degree of foreign accent whereas length of exposure and other factors played only a minor role. Moyer (1999) set out to challenge the CPH for phonology by testing highly motivated and experienced subjects on a variety of pronunciation tasks. Nevertheless, she found that non-native speaker performance was not within the range of native performance. Also Piske, Mackay, and Flege (2001) in their review on the existing literature on degree of foreign accent found that although sometimes other factors, such as motivation, L2 aptitude, and so forth also tended to have an impact, AoA was the most important predictor. However, also in this domain there are exceptions. Bongaerts, van Summeren, Planken, and Schils (1997) reported two studies where highly successful Dutch learners of English achieved an English accent which in their study proved to be indistinguishable from native speakers even though they had started to learn English only after the age of 12. Another two examples are provided in a study by Ioup, Boustagui, Tigi, and Moselle (1994) where two native English speakers acquired a native-like accent in Egyptian Arabic in adulthood.

(Morpho)syntax

There is ample evidence for greater difficulties with syntax and morphorsyntax with increasing AoA. For example, in the previously mentioned study by Johnson and Newport (1989), a decline in performance for L2 in a variety of tasks testing syntax (e.g., determiners, pronomilization, particle movement) and morphosyntax (past tense, plural and present progressive formation) was found for AoAs as early as 8 to 10 (p. 96). In the study by McDonald (2000) involving a GJT in L2 English testing the same rule types as Johnson and Newport but with early and late bilingual L1 Spanish and Vietnamese subjects, effects of AoA as well as native language influences were found[3]. Similarly, all of the studies reporting either CP or AoA effects cited above used some kind of a GJT. Also the earlier cited studies by Birdsong and Molis (2001) and Birdsong (1992) found age effects using the same material and procedure as Johnson and Newport. As mentioned above, Coppieters (1987) tested various aspects of syntax and morphosyntax, such as the usage of the two past tense forms imparfait vs. passé compose, the third person pronouns *il* and *ce* and pre-posed vs. post-posed adjectives and found evidence for CP/AoA effects. Patkowski (1980) evaluated knowledge of grammar, vocabulary and general communicative ability in oral interviews and Sorace (1993) tested unaccusativity in Italian, both finding evidence for CP/AoA effects.

Nonetheless, there were also studies that did not find a difference between native and L2 speakers in their grammatical judgment ability of, for example, the interpretation of tense aspect form in L2 Spanish (preterite-imperfect contrast) (Montrul & Slabakova, 2003) and *wh*-movement and question formation (White & Genesee, 1996). Moreover, the two native English-speaking subjects in the earlier cited study of Ioup et al. (1994) also performed mostly within the range of native speakers in a translation task and a (speeded) GJT in Egyptian Arabic. A wide range of syntactic constructs which were unique to Egyptian Arabic and not translatable to English, such as relative clauses, yes/no and *wh*-questions, word order in questions, conjoined NPs, and definiteness concord, to name just a few, were tested. On the other hand, further evidence for difficulties with L2 morphosyntax is provided in

[3] The studies by Johnson and Newport (1989), Birdsong and Molis (2001) and McDonald (2000) all used the same or very similar material testing the same rule types. They only differed regarding the language backgrounds of the L1 speakers (Korean and Chinese, Spanish and Vietnamese/Spanish, respectively). This raises the question whether the differential results (AoA effects, no AoA effects, AoA effects/no AoA effects, respectively) cannot rather be explained by L1 effects.

two more recent studies by Jiang (2004, 2007). The author conducted a self-paced reading study with Chinese L2 speakers of English and a native English control group (2004). He found that Chinese speakers were sensitive to verb subcategorization violations (∗*The teacher insisted the student to start all over again*) just as the native control group but not sensitive to number agreement violations (∗*The visitor took several of the rare coin in the cabinet*). The results were replicated later in a similar study (N. Jiang, 2007). So even if the behavioral evidence is somewhat mixed, it seems that late L2 learners often have problems with aspects of syntax and morphosyntax.

1.3.2 ERP studies

Further evidence for differences between L1 and late L2 speakers in the processing of syntax and morphosyntax is provided by neuroimaging studies. For the interpretation of ERP studies in this context, the ELAN component (early left anterior negativity) peaking around 150 - 200ms after stimulus onset and the P600 component (a positive deflection around 600 ms after stimulus onset) are especially relevant. The two components index grammatical processing and the detection of grammatical violations. They are thought to reflect automatic first-pass parsing processes (ELAN) and more conscious re-analysis (P600) (Mueller, Hahne, Fuji, & Friederici, 2005, p. 1230). In the context of morphosyntactic processing also the LAN (left anterior negativity) is important. The LAN reaches its maximum around 300 - 500 ms post-stimulus (Friederici, 2002). In semantic or world knowledge processing[4], the N400, a negativity with a centro-parietal distribution peaking around 400 ms, indicates the detection of anomalies. Below, I will first discuss the results of ERP studies investigating syntactic processing, followed by studies on morphosyntactic processing and lexico-semantic processing. Then I will take a look at studies investigating semantic processing and comparing lexico-semantic with (morpho)syntactic processing in L2 learners. As we shall see, there are important differences between semantic and (morpho)syntactic processing in L2.

Syntax

Some studies indicated that L2 learners, in contrast to L1 learners, showed no P600 effect when processing grammatical anomalies. For example, Hahne and Friederici (2001) conducted an experiment with native speakers of Japanese listening to German (L2) sentences which were either correct or contained semantic (*Der Vulkan wurde gegessen* = *The volcano was eaten*) or syntactic violations (**Das Eis wurde im gegessen* = **The ice cream was in-the eaten)*. The task consisted in judging the correctness of the sentences and accuracy was not excellent but well above chance. Different from the native control group, L2 learners showed no P600 effect (and no ELAN) in response to syntactic violations. So clearly, the Japanese speakers of German were not native-like in their processing of syntactic violations. Guo, Guo, Yan, Jiang, and Peng (2009) conducted a reading comprehension task and reported that native speakers showed a P600 in response to verb categorization violations (*Joe's father didn't let/*show him drive the car*), whereas L2 speakers exhibited an N400 instead. In other words, these studies provide examples of late L2 learners processing syntactic violations in a different way than native speakers.

[4] Note that as pointed out by Pylkkänen, Oliveri, and Smart (2009) there seems to be a "terminological difference between cognitive neuroscience and linguistics" (p. 2) regarding the distinction between semantic and world knowledge, which are used synonymously in cognitive neuroscience. Typical sentences eliciting an N400 are for example "While I was visiting my home town, I had lunch with several old shirts."(Luck, 2005. p. 45). However, this actually constitutes an example of a world knowledge violation because it is common knowledge that pieces of clothes are not able to enjoy a meal. However, in cognitive neuroscience this type of violations is usually called a "semantic violation". Since the present thesis cites a lot of studies from cognitive neuroscience, I will use the term "semantic violation" or "semantic processing" as used in the neurosciences.

In other studies, sometimes syntactic errors were processed similarly to native speakers by late L2 speakers as evidenced by the P600 component but differently regarding the ELAN component. Hence, L2 speakers were native-like in their re-analysis of syntactic errors but not regarding more automatic first-pass parsing processes. For example, Hahne (2001) conducted an auditory sentence judgment task with semantic as well as syntactic violations similar to the ones used in the previously described by Hahne and Friederici (2001). Subjects were Russian L2 speakers of German. Hahne (2001) found that sentences with a syntactic violation elicited an ELAN and a P600 in the native group, while only a P600 (which was slightly delayed) was observed in the L2 group[5]. So L1 and L2 processing was only similar regarding the P600 effect but not regarding the ELAN. Note, however, that ELANs are rarely found in L2 processing and even their occurrence in native processing is not uncontroversial (Müller & Hagoort, 2006). Mueller, Hahne, Fujii & Friederici (2005) also found a P600 in response to word category violations in a miniature version of Japanese for very high-proficient L2 speakers, while native Japanese speakers in addition to the P600 also exhibited an early negativity "similar to the ELAN" (p. 1238). Pakulak and Neville (2011) tried to disentangle effects of proficiency and AoA. They tested German L2 learners of English who were matched in proficiency to the native speaker group who belonged to the "lower-proficient" monolingual group of Pakulak and Neville (2010). They found a similar P600 for both groups (though with somewhat different temporal and spatial distribution) but no anterior negativity for L2 learners. Their conclusion was that AoA continues to play a role, even if proficiency is controlled for. Thus, even if in these studies a P600 was observed in response to syntactic violations also in late L2 learners, processing was still not native-like as the ELAN was only observed in the native control groups.

However, in the domain of syntax there are also cases where no differences between L1 and L2 speakers were found. Friederici, Steinhauer, and Pfeifer (2002) investigated the processing of an artificial language, called Brocanto, with highly trained participants by measuring ERPs. The stimulus material consisted of spoken sentences in Brocanto and half of the sentences contained severe phrase structure violations. The authors found that, in comparison to a control group that was only lexically trained, the highly trained experimental group processed sentences in a similar way as native speakers would do, namely, exhibiting an ELAN and a P600 in response to syntactic violations. The fact that this time L2 learners showed an ELAN, is taken as evidence against the CPH arguing that L2 language processing in adult learners can very well be based on the same brain mechanisms as an L1 and that L2 syntactic processing can be native-like. In a later study, an ELAN–P600 pattern in high-proficient L2 learners was also observed by Rossi et al. (2006) in response to word-category violations. Subjects were native speakers of Italian and German who spoke German and Italian as an L2, respectively, and a native German control group. Consequently, albeit native-like sensitivity to syntactic violations in L2 is by no means common, it is still possible. Further occurrences of native-like processing in late L2 speakers will be discussed in more detail in section 1.4 on the role of proficiency in L2 learning.

Morphosyntax

In the domain of morphosyntax, native-like processing is rarely found but not impossible. Osterhout, McLaughlin, Pitkänen, Frenck-Mestre, and Molinaro (2006) report an N400 instead of a P600 in very low-proficient learners in response to morphosyntactic violations (verbal person agreement and

[5] The fact that, contrary to Hahne and Friederici (2001), this time a P600 was found for the L2 group can probably be attributed to proficiency effects which will be explained in section 1.4.

number agreement). Ojima, Nakata, & Kakigi (2005) conducted a study investigating processing of subject–verb agreement errors by native speakers of Japanese with either low or high proficiency in their L2 English. They obtained a LAN in the high-proficient group but no P600, while both components were found in native speakers. Low-proficient subjects were completely insensitive to morphosyntactic violations. Also Chen, Shu, Liu, Zhao, and Li (2007) found a "biphasic LAN–P600 syntactic processing profile" (p. 171) for English native speakers in response to subject–verb agreement violations but not for native Chinese speakers proficient in their L2 English. In the previously cited study by Rossi et al. (2006), on the other hand, a native-like biphasic LAN–P600 pattern was found in response to subject–verb agreement (in addition to the ELAN–P600 pattern in response to syntactic violations). Their subjects were high-proficient L2 speakers of German and Italian who were native speakers of Italian and German, respectively. Hahne et al. (2006) investigated processing of regular and irregular past participles and noun plurals by high-proficient native Russians in their L2 German. L2 subjects showed an anterior negativity–P600 pattern in response to past participle violations but only a P600 (regular forms) and an N400 (irregular forms) in response to noun plural violations[6]. No native control group was used, but an anterior negativity had also been expected for noun plural violations. So, in the area of morphosyntactic processing the usual finding is that late L2 learners diverge from native speakers. The only exception was provided in the study by Rossi et al. where high-proficient learners performed native-like. As mentioned above, proficiency effects will be discussed in depth in the section 1.4.

Semantics

Regarding semantic processing, evidence for native-like processing in late L2 speakers abounds. In this context the ERP component of relevance is the N400 component. This component is observed in response to semantic violations (e.g., *I generally like menthol bottles.*) and therefore functions as an indicator of semantic processing. In an ERP study, Ardal, Donald, Meuter, Muldrew, and Luce (1990) found an N400 in response to semantic anomalies in monolinguals as well as high-proficient early and late bilinguals with the N400 of both bilingual groups differing only slightly from the N400 of the monolinguals. Furthermore, no AoA effects were found. Also Clahsen and Felser (2006) concluded in their review article on the native-likeness of non-native processing that N400 effects found for lexical-semantic processing are usually similar for native speakers and L2 learners, even if sometimes with slight differences in latency or amplitude. And in a PET study by Perani et al. (1998) it was demonstrated that native speakers and high-proficient late bilinguals showed activation in similar brain areas while listening to stories. Further evidence of native-like semantic processing in late L2 speakers has also been found in studies directly comparing (morpho)syntactic with semantic processing. These findings will be discussed in the following paragraphs.

Comparison of (morpho)syntactic and semantic processing

Especially useful for clarifying if different domains are really differently affected by AoA are studies directly comparing (morpho)syntactic and semantic processing in L2 speakers. ERP studies have revealed that AoA has differential effects on semantic and (morpho)syntactic processing, as evidenced by the N400 and P600 or ELAN component, respectively. These differential AoA effects have been observed for L1 as well as L2 processing. For L1, this was first observed in a reading experiment measuring ERPs by Neville, Mills, and Lawson (1992). Stimulus sentences contained no violations but ERPs were measured in response to open class words (semantic information) and

[6] The difference in processing regular and irregular past tense forms will be picked up again when treating the declarative/procedural Model (Ullman, 2001b) in the section on Models explaining AoA effects.

closed class words (syntactic information). Normal adults were compared to congenitally deaf adults with late exposure to English (upon school entry) and lower proficiency in English. The N400 component which was elicited by open class words was similar between the two groups, while ERPs elicited by closed class words (N280 and N400 - 700) were absent in deaf subjects (p. 255). Neville et al. (1992) concluded that the neural systems underlying semantic and syntactic processing are differentially constrained by different sensitive periods. In L2 processing, a similar processing difference between open and closed class words was found for late Chinese–English bilinguals by Weber-Fox and Neville (2001). (For a critique on this approach because of a confound with word length see Osterhout, Allen, and McLaughlin (2002)). The overall result, namely, stronger AoA effects for L2 syntactic processing than for semantic processing was confirmed by another ERP experiment by Weber-Fox and Neville (1996). They tested Chinese L2 speakers of English with five different AoA groups on a sentence judgment task involving semantic and syntactic violations (phrase structure, specificity constraint, subjacency constraint). All groups showed an N400 in response to semantic violations but ERP components found in response to syntactic violations differed depending on AoA[7].

Moreover, some of the previously cited studies on (morpho)syntactic L2 processing also investigated semantic processing. Similarly, the usual finding was that in terms of semantic processing L2 learners could not be distinguished from native speakers, while the processing pattern for syntactic violations looked different. In the aforementioned study with native speakers of Japanese who learned German as an L2, Hahne and Friederici (2001) demonstrated that in response to semantic violations in German the Japanese subjects exhibited an N400 as observed in native speakers, while in response to syntactic violations different to the native control group, no ELAN or P600 emerged. Similarly, Hahne (2001) found an N400 in response to semantic violations for native German speakers and Russian speakers of German as an L2, but only a P600 and no ELAN for the L2 speakers in response to syntactic violations. Sanders and Neville (2003) found processing differences between monolinguals and Japanese–English late bilinguals for syntactic processing but not for semantic processing. Ojima et al. (2005) measured ERPs of low- and high-proficient native Japanese speakers in their L2 English on sentences containing semantic and morphosyntactic violations. They found a native-like N400 in response to semantic violations (with slight time course differences), while, as mentioned above, processing of syntactic anomalies was not native-like. Ojima et al. (2005) concluded that language learning in adulthood resembles childhood learning in the sense that semantics is mastered before syntax. In an fMRI study, Wartenburger et al. (2003) found that the cortical representations for early and late bilinguals differed only in the case of morphosyntactic processing (subject–verb and gender agreement). The pattern of brain activity for semantic judgment, though, largely depended on proficiency with brain areas of early and late bilinguals overlapping more when the late bilingual is high-proficient. Thus, the cortical representations of grammatical processes are more affected by AoA than proficiency. So recent research seems to point into the direction that late bilinguals have few problems with semantic processing, whereas regarding syntactic and morphosyntactic processing results are more mixed.

Hence, there are great differences in AoA effects for the different domains. AoA effects seem to be strongest for phonology. But also syntax and morphosyntax are difficult for adult L2 learners and native-like attainment is rarely found. In the domain of semantics, however, adult L2 speakers are

[7] The N400 in response to semantic violations for the two “older” AoA groups (11-13 years and > 16 years) was also slightly different from the “younger” groups.

often indistinguishable from native. This differential effect of AoA on (morpho)syntactic and semantic processing becomes especially apparent in studies directly comparing these two domains. Yet, the cause for this difficulty is still unclear. For this reason, the focus of the present thesis is on an aspect of morphosyntax that is very difficult to acquire, namely, grammatical gender. (Findings on L2 processing of grammatical gender will be discussed in chapter 2 and section 3.3). Because of the problems grammatical gender causes in L2 learning, it can probably give informative insights on differences between L1 and L2 acquisition and processing. In the next section, I will take a closer look at the role of proficiency in adult L2 processing, which is another central topic of the present thesis. Studies providing evidence of native-like attainment, also in the domains of syntax and morphosyntax, will be discussed. In addition, I will report studies investigating changes in L2 processing with increasing proficiency and in the course of L2 learning.

1.4 The role of proficiency

As described earlier, AoA seems to be the most important predictor for success in L2 acquisition. However, as was implied in some of the studies summarized before, this is not a matter of “all or nothing”. Rather, there is evidence that very high-proficient late L2 learners can in some cases make up for AoA effects and attain native-like proficiency and native-like processing patterns, as we have seen especially in the domain of semantics. In fact, Newman, Tremblay, Nichols, Neville, and Ullman (2012)

conducted a study investigating proficiency effects on ERPs in response to semantic violations in native English speakers and Spanish L2 speakers of English. Based on their results, they argued that the greater N400 amplitude oftentimes found for native speakers (Hahne et al., 2006; Moreno & Kutas, 2005; Ojima et al., 2005; Weber-Fox& Neville, 1996) is mostly influenced by proficiency, while only latency and distribution are influenced by AoA. In two PET studies, Perani et al. (1996, 1998) showed that semantic processing of L2 stories depended more on L2 proficiency than on AoA. In addition, the previously cited study by Wartenburger et al. (2003) showed that in the case of semantic processing proficiency was the most important predictor for native-like processing, while in the case of morphosyntactic processing it was AoA. But also in the domain of (morpho)syntax which, as discussed above, is especially prone to AoA effects, proficiency plays an important role. Next, I will discuss studies that found evidence for high-proficient late L2 learners performing like natives even in the domain of (morpho)syntax.

1.4.1 High-proficient late L2 learners perform like natives

In recent years, the factor of L2 proficiency next to mere AoA effects has gained increasing importance in research. Perani and Abutalebi (2005), for example, point out the importance of proficiency and amount of L2 exposure next to AoA in shaping L2 brain representations and levels of brain activation. They discuss recent evidence for the representation of L1 and L2 in the same brain areas even in cases of late acquisition. They claim that differences between native and L2 speakers disappear with increasing proficiency. Furthermore, in their discussion reviewing various ERP results on AoA and proficiency effects in the domain of morphosyntax, Steinhauer, White, and Drury (2009) concluded that proficiency is more important than AoA. This is because proficiency seems to be a better predictor of brain activity patterns than AoA and native-like processing patterns are possible at near-native proficiency levels. The authors do not support the assumption of Clahsen and Felser (2006) that some linguistic structures cannot be acquired by late L2 learners. They rather put an emphasis on the fact that morphosyntactic processing also changes in the course of L2 acquisition, just as semantic processing.

Some studies providing evidence for native-like attainment in high-proficient late L2 learners have already been mentioned in the previous section. For example, in the aforementioned study, White and Genesee (1996, cf. section 1.2) concluded that even late L2 learners can achieve native-like proficiency. They carefully selected high-proficient early and late L2 speakers of English with Germanic and Romance language backgrounds and divided them into a non-native and a near-native group after evaluation of picture-elicited spontaneous speech samples. Only subjects with speech samples indistinguishable from the native control group were chosen for the near-native group. Their results showed that these near-native subjects performed as well as native speakers on a GJT and a question formation task in terms of speed and accuracy measures, therefore showing similar competence as well as performance. Furthermore, also the aforementioned studies by Birdsong (1992) and Montrul and Slabakova (2003) found cases of native-like attainment among late L2 speakers. Another example is provided by the likewise previously cited study of Ioup et al. (1994) where two native English-speaking women with AoAs in adulthood performed within native-speaker range in various grammatical tasks in Egyptian Arabic. Especially the native-like performance of Montrul and Slabakova´s (2003) and Ioup et al.´s (1994) subjects is impressive as tense aspect form in Spanish and the Arabic grammar rules tested are supposed to be very difficult to master for (English) L2 speakers (as stated by the authors).

Moreover, native-like performance of highly-proficient late L2 speakers has not only been demonstrated in behavioral studies. Also in ERP studies native-like patterns can be observed. The previously cited study by Rossi et al. (2006) found that high-proficient speakers of L2 German and L2 Italian exhibited the same components (ELAN/LAN and P600) to all types of (morphosyntactic) violations as a control group of native speakers of German. Except for some amplitude differences, processing was identical to native speakers´. Low-proficient speakers, however, showed fundamental differences. Rossi and colleagues (2006) take that as evidence that native-like processing can be attained by late learners with high proficiency levels.

In addition, the different components (P600, ELAN) that were found for L2 speakers in the aforementioned studies by Hahne and Friederici (2001), Hahne (2001), and Friederici et al. (2002) can possibly largely be attributed to proficiency effects. The Japanese speakers of German in the study by Hahne and Friederici (2001) were probably the lowest proficient (mean self-rated proficiency 3.5 out of 6, grammatical judgment accuracy between 66 and 86 % across conditions) and showed no P600 effect in response to grammatical violations. The more proficient Russian group (mean self-ratings between 2.9 and 3.7 out of 4 for different language skills, accuracy between 92 and 93 % across conditions) in the study by Hahne (2001), however, showed a P600 but no ELAN in response to grammatical violations. It is possible that the subjects of the artificial language study by Friederici et al. (2002) were in turn even more proficient, at least with respect to the limited scope of structures tested. Subjects were extensively trained in syntax (until 95 % accuracy was reached) and only had to memorize a small vocabulary of 14 words. Grammatical judgment accuracy was high (93 %) and in addition to a P600 subjects also showed an ELAN in response to grammatical violations, a component that is rarely observed in late L2 learners (cf. section 1.3.2). In another study by Hahne et al. (2006), L1 Russian–L2 German speakers exhibited an anterior negativity–P600 pattern in response to past participle violations but no anterior negativity was found in response to noun plural violations in the L2. It appears that noun plurals were more difficult for the L2 subjects than past participles, as indicated by error rates in an elicited production task. Or, put differently, the L2 subjects had a higher proficiency for past participles than for noun plurals which explains the lack of an ELAN for the latter. Thus, even in late learners proficiency effects are important and L2 processing does not have to remain different from native processing due to AoA effects. This invites the question how L2 processing may change in the course of L2 learning and with increasing proficiency. In the next section, I will look at studies that investigated how L2 processing changes in the course of L2 learning. This issue is important for the present thesis as across the two experiments, the performance of L2 learners of very different proficiency levels, from very low to very high-proficient, will be investigated.

1.4.2 Changes in the course of L2 acquisition

Quite a few longitudinal studies focusing on processing changes throughout the course of L2 acquisition in the different language domains have been conducted. In the domain of semantics, McLaughlin et al. (2004), for example, conducted an ERP study with a primed LDT with native English speakers of L2 French and showed that after only 14 hours of classroom instruction, French pseudowords elicited a larger N400 than real words. Effects of prime type (semantically related or unrelated) were seen after 60 hours of instruction, with smaller N400s in response to targets preceded by a related prime than an unrelated prime. After 140 hours, the N400 pattern was almost native-like across all conditions. A similar result was obtained in an experiment reported by Osterhout et al. (2006). Here, the L2 learners of French (also L1 English) exhibited an N400 after only 1 month of instruction in response to semantic violations in a sentence processing task. This shows that within the domain of semantics, rapid changes in L2 processing are possible after only a short period of instruction.

Moreover, also in the domain of (morpho)syntax, L2 processing can develop within short time periods. Osterhout et al. (2006) report a study that showed that after 1 month of instruction, their L2 learners exhibited an N400 in response to verbal person agreement violations instead of a P600. After 4 months, however, L2 subjects exhibited a “P600-like” positivity (p. 219), indicating sensitivity to this type of morphosyntactic violations. This fast shift from non-native-like processing to exhibiting a component observed in native processing is impressive. Nevertheless, processing changes have been observed after even shorter time periods. Davidson and Indefrey (2009), for example, showed that native Dutch speakers of German as an L2 were able to develop sensitivity to morphosyntactic declension violations as evidenced by a P600-like response after only a short training phase provided a week before. Also the earlier cited studies of Friederici et al. (2002) and Mueller et al. (2005) involving learning of an artificial language and of a miniature version of Japanese, respectively, report native-like P600s (and even an ELAN in Friederici et al.) in response to L2 syntactic word category violations (Friederici et al. (2002) and Mueller et al. (2005)) and morphosyntactic case violations (Mueller et al.). This is noteworthy because acquisition of these artificial/miniature languages occurs in training sessions within a very limited period of time (“[…] several training sessions […] up to 5 h per session” in the study by Friederici et al. (p. 530) and 4 to 10 hours in the study by Mueller et al.). So even if, as discussed in the beginning of the chapter, high AoA seems to impair L2 processing, it is certainly impressive how fast processing changes in the brain can be observed after a limited amount of L2 exposure.

Hence, proficiency plays an important role in L2 processing. It can be responsible for changes in the brain resulting in native-like processing patterns as observed in ERPs and thus compensate for negative AoA effects. Nevertheless, it is important to note that differences in proficiency are by no means exclusively observed in L2 learners. Proficiency differences have also been found in native speakers. For example, Pakulak and Neville (2010) measured ERPs of native English speakers of different proficiencies (from different social classes and educational backgrounds) in response to aurally presented phrase structure violations. Their results showed that both proficiency groups (high and low) exhibited a LAN and a P600, but the components differed slightly in topography and temporal distribution (LAN) or amplitude (P600). Weber-Fox, Davis, and Cuadrado (2003) investigated processing differences between normal and high-proficient native speakers of English as indexed by ERPs and found differences for the P200[8] and later components. Malaia, Wilbur, and Weber-Fox (2009) examined processing of telic and atelic garden path sentences in normal and high-proficient native speakers of English. They observed that amplitude differences between conditions occurred at an earlier time point in the sentence for high-proficient than for low-proficient subjects. As mentioned earlier, Newman, Tremblay, Nichols, Neville, and Ullman (2012) found larger N400 amplitudes in response to semantically appropriate words for lower-proficient native speakers as well as L2 speakers. In short, proficiency differences are not only observed between native and late L2 speakers, but also between groups of native speakers.

Furthermore, there are many other factors that influence the processing of L2 grammar. For example, working memory seems to play an important role (McDonald, 2006; Perani, 2005; Sagarra & Herschensohn, 2011a), L1 and L2 characteristics and similarity (Frenck-Mestre, Osterhout, McLaughlin, & Foucart, 2008; McLaughlin, Tanner, Frenck-Mestre, Valentine, & Osterhout, 2010; Tokowicz & MacWhinney, 2005; cf. also sections 3.3.1 and 3.3.2), as well as sociolinguistic factors (Ortega, 2009, chapter 10), or personality factors such as extraversion (Dewaele & Furnham, 1999, 2000). This thesis focuses on proficiency effects in the context of grammatical gender acquisition, also touching on the subject of L1 and L2 characteristics and similarity of their grammatical gender systems.

As we have seen, evidence for proficiency effects is numerous. This is true both for differences between low- and high-proficient L2 speakers and for high-proficient speakers approaching native-likeness. In some cases, high-proficient late L2 speakers are able to make up for AoA effects and perform as well as early L2 speakers or native monolinguals, even on grammatical tasks. Proficiency effects have been observed in behavioral as well as ERP tasks, showing that late L2 speakers can even be indistinguishable from early L2 speakers or monolinguals in online-measures. In other studies it has been shown that there are different stages in the course of L2 learning before an L2 learner eventually reaches native-like proficiency in a certain area. Changes at the neurological level already occur after short exposure to an L2 within a few hours of classroom instruction. This suggests that processing patterns of adult L2 learners are not rigid and unchangeable, but on the contrary are highly sensitive to L2 input. Therefore, to gain a better understanding of late L2 acquisition it is necessary to further study proficiency effects.

In addition to the proficiency effects discussed here, in previous sections, we have also learned a lot about the importance of age effects in language learning and some of the aspects related to them. Regarding age effects, several questions remain to be answered. For example, what is the reason for

[8] A positive deflection found around 200 ms after stimulus onset is denoted a P200. Note that even though the P200 is oftentimes observed in language processing, it is typically not reported as its exact interpretation in language processing is still unclear. It has more commonly been reported in studies investigating memory and perception (cf. section 5.4.2 for a broader discussion of the P200 component).

those age effects? And why and how can these age effects be mediated by proficiency? In the next section, I will present two models that attempt to provide an explanation for AoA effects in language learning. Both models have incorporated proficiency, which underlines the importance of language competence next to age effects. Even if none of the models is able to provide an explanation for all of the aspects important for the present thesis, namely, the difficulty of grammatical (gender) processing, proficiency effects and language transfer in late L2 learners, each model is able to account for some of them.

1.5 Models explaining AoA effects

Hence, what is the reason for the differences in L2 attainment between younger and older learners and the differential findings for semantic and syntactic processing? Besides the CPH discussed earlier in this chapter, a great deal of theories and models trying to explain AoA effects in L2 learning have been developed. For the domain of grammar, the most well-known theory is certainly the Universal Grammar theory (applied to L2 acquisition by White, 1989) on which the Failed Functional Features Hypothesis (Hawkins & Chan, 1997) and the Full-Transfer Full-Access Model (Schwartz & Sprouse, 1996) are based. Another explanatory attempt is provided by the Shallow Structure Hypothesis (Clahsen & Felser, 2006a, 2006b, 2006c). For the present thesis, the Declarative/Procedural Model (DP Model) by Ullman (2001a, 2001b, 2001c, 2004) and the Competition Model by MacWhinney (1987a, 1987b, 2005a) are especially relevant. The DP Model provides an explanation for AoA effects and the differential effects on vocabulary and grammar, as well as for proficiency effects (cf. previous section). The Competition Model provides a framework within which AoA effects as well as L1 transfer and proficiency effects can be described.

1.5.1 The Declarative/Procedural Model

The DP Model was originally developed to explain L1 use (Ullman, 2001a) and was then extended to explain L2 learning and use (Ullman, 2001b). Concerning L1, Ullman departs from the assumption that in language processing, mainly two well-known memory systems are involved: the explicit/declarative memory system, which is located in temporal lobe structures, and the implicit/procedural memory system located in left-frontal/basal-ganglia structures. The declarative memory system underlies the learning and processing of vocabulary, whereas grammar is acquired and processed relying on the procedural memory system. Moreover, information stored in the declarative memory system is explicit and conscious, while the knowledge of the procedural system is implicit and automatic.

This approach has some commonalities with what Ullman (2001a) calls “traditional” (p. 107) “single-mechanism” theories (mostly connectionist models, e.g., MacWhinney & Leinbach, 1991; McClelland & Rumelhart, 1985) and foremost “dual-mechanism” theories (e.g., Paradis, 2004), but there are also differences. Single-mechanism theories claim only one broadly-distributed associative memory system for the storage of vocabulary and the storage of descriptive grammatical rules together, whereas “traditional” dual-mechanism theories also postulate two separate systems underlying the acquisition and processing of vocabulary and grammar. However, contrary to traditional dual-mechanisms theories, in the DP Model these systems are not posited to be exclusively dedicated to language. The declarative memory system contains not only information specific to language but also other facts from semantic and episodic memory. The procedural memory system is also involved in the learning of complex motor skills and habits. Furthermore, under traditional dual-mechanism accounts, lexical memory is thought to be just a rote list of words, while under the DP account lexical memory possesses the characteristics of semantic and episodic memory, namely, that facts are structured and associated with each other. Next to the grammar-lexicon distinction, the DP Model makes predictions concerning the processing of regular vs. irregular items.

As illustrated by the example of English past tense morphology (e.g., Pinker & Ullman, 2002; Ullman, 2001c), according to the DP Model, the declarative and procedural memory system interact in such a way that regular forms (*-ed* suffix) are computed by the "rule-system", the procedural memory system, while irregular forms (e.g., *go-went*) are stored and retrieved by the declarative memory system[9]. When an irregular past tense form is encountered in declarative memory, the computation of the regular form is inhibited. In this sense, the DP Model appears to be something of a combination of the two traditional approaches. In the DP Model, the associative lexical memory still does not compute rules, but it has the ability to extract, learn and apply patterns to new forms (Ullman, 2001c, p. 42) as in the case of, for example, *grind–ground*, *find–found*, *wind–wound*. Evidence for this distinction between regular and irregular forms is provided, for example, by Bowden, Gelfand, Sanz, and Ullman (2010) and in an ERP study by Newman, Ullman, Pancheva, Waligura, and Neville (2007)[10]. Further evidence regarding regular/irregular English verbs (e.g., Prado & Ullman, 2009; Ullman, 1999; van der Lely & Ullman, 2001) and regarding regular/irregular German noun plurals (Clahsen, Eisenbeiss, & Sonnenstuhl-Henning, 1997; Penke & Krause, 2002) is cited by Ullman (2001a). Moreover, due to these different functional specializations of the two memory systems, neurological dissociations for morphological transformations are predicted by the DP Model[11]. Further evidence from studies on aphasia, neurodegenerative diseases, and developmental disorders is cited in support of this double dissociation[12] (e.g., Clahsen & Almazan, 1998; Ullman & Gopnik, 1999; Ullman et al., 1997, 2005).

The DP Model was also extended to explain L2 learning and AoA effects in late L2 learning, especially the differential effects on semantic and syntactic processing (Ullman, 2001b) discussed above. The model builds on findings that learning through procedural memory is at its best during childhood and slowly deteriorates thereafter (Morgan-Short & Ullman, 2011). Due to these maturational constraints, late L2 learners come to rely more heavily on declarative than procedural structures, which leads to good vocabulary acquisition but imperfect grammar acquisition. However, with practice, augmenting proficiency and automation of the L2, it is also possible for late L2 learners to increasingly rely on the procedural memory system again. Thus, the DP Model does not preclude success in L2 acquisition for adult learners but explains the mechanisms at work when late bilinguals overcome age effects. In the previous section, we looked at the role of proficiency and how in some cases late bilinguals can become high-proficient compensating for AoA effects. The model also predicts greater difficulties for late L2 learners at greater agreement distances, because "local

[9] But note that the likelihood for even regular forms to be stored in associative memory increases with the frequency of the form (Ullman, 2001c, p. 43) and might also differ between sexes (Prado & Ullman, 2009).

[10] Note, however, that the DP Model has also been criticized amongst others things for the fact that it has been mostly investigated and discussed with respect to its application to inflectional morphology, especially English past tense (Embick & Marantz, 2005; MacWhinney, 2005a; but see Ullman & Walenski, 2005 for a reply).

[11] These predicted dissociations are different than under a traditional dual-mechanism account (Ullman, 2001a). A detailed analysis of the differences between traditional dual-mechanism accounts and the DP Model is beyond the scope of this section, though.

[12] Most of these studies also provide evidence for the postulated localizations for the two memory systems specified by the DP Model, but this evidence will not be considered here as matters of neuroanatomical localization are of less importance for the present thesis.

dependencies should be easier to learn (e.g., as chunks) than nonlocal dependencies in lexical/declarative memory" (Morgan-Short, Sanz, Steinhauer, & Ullman, 2010, p. 182). The difficulties of late bilinguals at greater syntactic distances, especially with respect to grammatical gender, will be discussed in section 2.2.

There is also experimental evidence providing support for the predictions concerning AoA effects made by the model. Some of the ERP studies which obtained differential results for semantic and syntactic acquisition in late L2 learners summarized in section 1.3.2 can be taken as evidence in favor of the DP Model (Ullman, 2012). These studies found native-like semantic processing patterns but not native-like syntactic processing patterns for late L2 learners (Hahne & Friederici, 2001; Hahne, 2001; Neville et al., 1992; Sanders & Neville, 2003; Weber-Fox & Neville, 2001), some of which are also cited by Ullman in support of his model. Also the earlier mentioned study by Wartenburger et al. (2003) provided evidence for separately located memory systems for grammatical and semantic knowledge.

Furthermore, there is evidence from studies showing that sometimes at beginning levels subjects show an N400 in response to grammatical violations instead of a P600 (Guo, Guo, Yan, Jiang, & Peng, 2009; Morgan-Short, Steinhauer, Sanz, & Ullman, 2012; Osterhout, McLaughlin, Pitkänen, Frenck-Mestre, & Molinaro, 2006, experiment 2). The P600 is an index of true syntactic processing, while N400 is an index of semantic processing thought to rely on declarative memory structures (Morgan-Short, Steinhauer, et al., 2012). This supports the notion that at least in the beginning, late language learners use the declarative instead of the procedural system to process L2 grammatical structures. In their longitudinal studies, Osterhout et al. and Morgan-Short et al. provided evidence for the fact that this misapplication of the declarative system is overcome with greater proficiency when subjects start to show a P600 to grammatical violations, indexing correct grammatical processing.

As mentioned before, the DP Model has primarily been specified for regular and irregular forms of inflectional morphology, especially English past tense. Next to noun plurals, this has also been the main area of research. The most conclusive results regarding L2 processing of inflectional morphology are provided in the previously cited ERP study involving L2 German past participles and noun plurals by Hahne, Mueller, and Clahsen (2006). In a reading task they obtained N400s for "overirregularizations", that is, application of irregular suffixes to regular items and P600s for "overregularizations", that is, application of regular suffixes to irregular items for both past participles and noun plurals. In principle, this pattern indicates storage for irregular items and computations for regular items as in native speakers. However, LANs were only obtained for overregularizations of past participles but not of noun plurals formation. German noun plurals underlie more complex rules and subjects made more mistakes in noun plurals than in past participles. Therefore, the authors concluded that noun plural processing was probably still less automatized than past participle formation. This study shows that with high proficiency, L2 processing can resemble L1 processing.

Important for the present thesis, also regarding research on grammatical gender a few researchers have tested some of the assumptions made by the DP Model. In a study by Kempe, Brooks, and Kharkhurin (2010) it was found that after training sessions adult L1 English subjects had problems generalizing gender categories from Russian diminutives to simplex nouns. The authors concluded that, in line with the DP Model, it is difficult for adult L2 learners to detect and use regularities as they instead seem to rely on rote strategies. This is similar to the performance of L2 learners of Dutch in a study by Sabourin, Stowe, and De Haan (2006). They showed great accuracy (above 80 %) when assigning gender to Dutch nouns and were thus apparently able to recall gender lexically, potentially relying on the declarative memory system. But when having to perform morphosyntactic agreement between a noun and a relative pronoun, according to the DP Model a procedural memory task, accuracy dropped significantly. (More studies investigating the effects of agreement distance and thus lexical and morphosyntactic processes in gender processing will be discussed in section 2.2). The predictions made concerning the dissociation between a rule-based system and a lexically-based memory system sensitive to frequency effects were investigated with respect to grammatical gender by Blom, Polisenska, and Weerman (2008). They investigated gender errors made in L1/L2 Dutch by L1 Dutch children, L1 Moroccan Arabic children, and L1 Moroccan Arabic adults. Error patterns between article-noun and adjective-noun constellations were expected to be consistent for children, using a syntactic strategy but not for adults, using an input-based, lexical strategy. Based on the error patterns of the different groups in article-noun and attributive adjective-noun constellations, they were able to conclude children used a syntactically based strategy in L1/L2 Dutch. That is, gender errors made within the determiner phrase were also made in adjective-noun constellations and vice versa. The error profiles of adults, on the other hand, were inconsistent between the two structures tested. They also revealed clear effects of input frequency, indicating that adults used a more lexicon-driven strategy. This supports the notion of impaired access to the procedural learning route in adults which is replaced by a strategy relying on declarative mechanisms as postulated by the DP Model.

In sum, the DP Model is able to provide a convincing explanation for AoA effects in L2 learning, especially the finding that, as discussed in section 1.3 different domains are differently affected by age effects. Also the role of proficiency in L2 learning is accounted for. However, it is not clear what the predictions concerning L1 transfer effects and L1–L2 similarity would be. These effects in L2 learning will be discussed in chapter 3. Furthermore, in section 2.2, I will discuss the importance of task demands (performance in online/offline tasks) and agreement distance in L2 grammatical gender processing. In the following section, the Competition Model will be discussed. Important for the present thesis, the Competition Model is not only able to account for L1 transfer effects and L1–L2 similarities but language transfer, or language competition, is at the very core of the model.

1.5.2 The Competition Model

The Competition Model was developed within the framework of emergentism. The greatest difference between the Competition Model and the DP Model certainly is that the Competition Model is a single-mechanism model and proposes that L1 and L2 are processed by the same underlying mechanisms. Similar to the DP Model, the Competition Model was first developed to explain L1 (end-state) use (E. A. Bates & MacWhinney, 1987; MacWhinney, 1987a), L2 (end-state) use (MacWhinney, 1987b) and then extended to provide a better account for the L1 and L2 learning processes (MacWhinney, 2005a, 2005b). MacWhinney (2005b) underlines the importance of having just one model that accounts for L1 as well as L2 by stating that “the fact that L2 learning is so heavily influenced by transfer from L1 means that it would be impossible to construct a model of L2 learning that did not take into account the structure of L1” (p. 70). I will first explain the general mechanisms that operate according to the Competition Model and then describe the mechanisms that are important in L2 learning. In my discussion of the Competition Model I will mostly rely on its more recent version (MacWhinney, 2005a, 2005b).

The extended or unified version of the Competition Model, which provides a single framework for L1 and L2 acquisition and use, consists of seven components. To the original components of competition, arenas, cues, and storage, the components of chunking, codes, and resonance[13] were added (MacWhinney, 2005a):

- *Competition and arenas:* Competition is still at the core of the model and is modified by cue strength and resonance, which is similar to interactive activation in earlier versions of the model. Competition means that, for example, in the process of language production or comprehension, various candidates compete for selection at all times. As all languages known by a speaker are constantly activated at least to a certain degree, competition also occurs between languages. Competition arises in four linguistic arenas, namely, phonology, lexicon, morphosyntax, and conceptualization. The arenas of lexicon and morphosyntax are especially relevant for the processing of grammatical gender. In production, competition likewise occurs at four points, namely, message formulation, lexical activation, morphosyntactic arrangement and articulatory planning. The arenas of lexical activation and morphosyntactic arrangement are also especially relevant for grammatical gender processing. In comprehension, competition occurs in auditory processing, lexical activation, grammatical role decoding (relevant for grammatical gender), and meaningful interpretation. Importantly, the different arenas of a language do not constitute encapsulated modules but interact.

- *Cues:* The notion of cues indicates the importance of the mappings between forms and functions. A cue can be seen as a type of clue to, for example, the grammatical function of a word, such as inflectional or other suffixes. That way, the *-a* ending of a Spanish noun indicates that the noun most likely is feminine, while other endings are quite ambiguous as to the gender of the noun. Ambiguity plays an important role in cue strength. Cue strength, that is, the strength of the form–function mapping, is determined by cue validity, which is a function of cue availability and reliability. Frequent and reliable cues (the *-a* ending in Spanish nouns) are more valid and stronger than infrequent and ambiguous cues. This will be important for Experiment 1 where gender transparency is investigated. The stronger a cue is, the more likely it is that it wins the competition for selection over another cue. Other cues constitute for example the *-s* ending in English present tense verbs, which indicates third person singular. The rest of the present tense endings of English verbs, on the other hand, is quite ambiguous as to which person reading is required. Cues differ across languages. A cue to the object of a sentence can, for example, be word order or certain case markings, depending on the language.

- *Storage, chunking, resonance:* These form–function mappings, that is, the meaning of cues, must be stored in memory. That way, the learning and processing of mappings is constrained by the component of storage. Memory capacity has an influence on perceived cue validity. Only cues that can be processed and stored are analyzed in order to extract the nature of the mapping. The size of the mappings that are processed and stored, again, depends on the process of chunking. Single words or whole phrases can be stored as separate chunks. The component of codes describes the relative activation of the languages known by the individual, which can result in language transfer in the different linguistic arenas. Through the process of code interaction, languages are selected, switched, or mixed. Language activation is determined by resonance. Finally, resonance can be

[13] Note that in MacWhinney´s (2005b) article the seven components are named slightly differently, namely, competition, arenas, mappings, chunking, storage, codes, and, support. In the present discussion of the model I will adhere to the names used by MacWhinney (2005a).

thought of as a type of interactive activation within languages or arenas and is also important for the learning of new items and their successful storage in long-term memory. As in an interactive activation network, the activation of one node spreads to other nodes related to it. So when a word of a certain language is activated, it is likely that other words in the same language are also activated. But chances are that at least a few related words in another language will also become activated. The processes of storage, chunking, and resonance are involved in learning.

Important for the present thesis, the Competition Model specifies different mechanisms that are responsible for AoA effects in L2 learning. This is done without the necessity of postulating a CP for language learning. These principles which underlie emergentist theory in general also apply to the Competition Model[14] (Hernandez, Li, & MacWhinney, 2005). The mechanisms are competition, resonance, parasitism, and entrenchment. In the context of bilingualism, competition means above all that, when processing a language, all languages known to the speaker are activated and therefore compete with each other. Through the principle of resonance, at least in early bilinguals the needed language wins in this competition in most cases. This is because it receives most activation (resonance) through other words in the sentence in the same language as there are more connections between words and concepts in the same language. Late L2 learners have problems in L2 processing because of parasitism, which means that their L2 vocabulary and processing relies a lot on L1 vocabulary and L1 mechanisms. This is due to greater L1 entrenchment than, for example, in child L2 learners. Entrenchment is, as one could say, the degree of consolidation of a language and increases with experience. Greater L1 use leads to greater L1 entrenchment. As L2 speakers become more fluent, parasitism can decrease to a certain extent as a result of increased resonance and L2 entrenchment. The Competition Model also does not preclude native-like proficiency for late L2 learners. High L2 proficiency is possible through greater L2 entrenchment and resonance achieved by extensive practice. The Competition Model also postulates that entrenchment is not equally strong for the arenas of a language, being the strongest for phonology and the weakest for the lexicon (MacWhinney, 2005a, p. 63). This is consistent with the difference in AoA effects found for the different language domains, as discussed in the previous section[15].

Naturally, experimental evidence is provided for the different mechanisms thought to be at work in L2 processing. The claims made by the Competition Model, especially concerning competition and cue strength, have been validated in numerous studies in 18 different languages, involving children and adults, monolinguals as well as bilinguals (MacWhinney, 2011, p. 213; cf. also Su, 2001, p. 85). Experiments and corpora studies are carried out by putting different cues from L1 or L1 and L2 into competition with each other. By orthogonal variation of cues reliable data on the workings of cue competition in sentence comprehension are obtained.

The Competition Model further predicts that adult L2 learners have to additionally rely on non-language areas in the brain when processing L2 in order to increase L2 resonance and to compensate for L1 entrenchment. Similar to the DP Model, the Competition Model likewise assumes that children rely more on implicit processes than adult learners, who also have to employ non-language areas in

[14] In MacWhinney (2011 and 2008) some of these and the earlier discussed factors together with other factors (e.g., social factors such as “isolation” and “participation”) are described as risk and support factors for L1 learning. The general idea remains the same, though.

[15] Note that, however, the reason for this differential entrenchment across language arenas is not further specified.

L2 processing in order to increase L2 resonance and to overcome L1 entrenchment (Hernandez, Li, MacWhinney 2005, p. 4). This is in line with evidence from neuroimaging studies which have repeatedly shown that late L2 learners, especially with lower proficiencies, show more pervasive activity than early bilinguals or monolinguals also in non-language areas (Perani et al., 1996, 1998; Wartenburger et al., 2003). Hernandez et al. (2005) and Hernandez and Li (2007), on the other hand, cite studies which investigated processing of their L1 Korean in children who were adopted by French families before age 9 (Pallier et al., 2003; Ventureyra, Pallier, & Yoo, 2004). Findings showed that these children who obtained native-like proficiency in French showed native-like neural activity when processing French but their neural responses to Korean stimuli were indistinguishable from neural activity patterns to other unknown languages. Interpreted in terms of the Competition Model, this would mean that, since the Korean children were still young when they moved to France, their L1 was not fully entrenched, therefore allowing them to learn their L2 without being parasitic on L1. With time, L1 resonance further decayed so that the L2 could be processed without L1 competition.

One more specific prediction of the Competition Model is that structures that are similar in L1 and L2 or unique[16] to L2 will be easier to acquire than structures that differ in L1 and L2. In the case of similar structures, L2 speakers can profit from positive transfer effects from the L1. Unique structures are easier than different structures because no competition can occur. Different structures, on the other hand, are thought to provide problems because of competition from the L1. This is important, for example, in the case of native English speakers trying to acquire an L2 gender system. The predictions regarding similar, unique, and different structures were tested by Tokowicz and MacWhinney (2005). Low-proficient L1 English speakers in their L2 Spanish were tested on a GJT while also measuring ERPs. Tense marking was used as a structure that is similar in L1 and L2 and determiner gender agreement as a structure that is unique in L2. Determiner–number agreement is a structure that differs between the two languages and should therefore be the most difficult for the L2 subjects. A P600 for violations to the similar and unique structures but not to the “differential” structure was obtained. Thus, as predicted by the Competition Model, subjects only had acquired the two former structures but showed no sensitivity to violation of the latter. Furthermore, the Competition Model predicts that ERP effects should be stronger for structures that are similar in both languages because of cue summation due to L1 parasitism. This prediction was also borne out by the results. Curiously, behavioral measures were at chance for all three structures[17].

Interesting for the present thesis is that transfer is a very important aspect in the Competition Model because of the central role of competition. Transfer means that in a bilingual or multilingual, other languages than the target language are activated and exert an influence, which can be positive or negative, in the processing of the target language. (Language transfer will be discussed in depth in chapter 3). In fact, MacWhinney (2005a) even stated that “The basic claim is that whatever can transfer will”[18] (p. 55). When structures are similar, positive transfer occurs, when structures are different, negative transfer occurs (MacWhinney, 2011, p. 220). The Competition Model predicts L1

[16] MacWhinney (2011) acknowledges that unique L2 structures are not necessarily easy to acquire. They are free from negative transfer and thus competition but also lack the support of positive transfer and have to be acquired “from the bottom up without any support from the L1” (p. 220).

[17] ERP analyses included correct as well as incorrect trials.

[18] This is contradictory to Pienemann, Di Biase, Kawaguchi, and Håkansson´s (2005) claim that “only those linguistic forms that the learner can process can be transferred to L2". This issue is treated by MacWhinney (2008, pp. 351-352).

transfer in all linguistic arenas due to L1 entrenchment and competition between all languages known to a speaker. Regarding syntax, MacWhinney (2005a) explains that L2 learners first interpret L2 cues as they would in an L1 sentence. Only gradually the cue weight settings start approximating those of an L2 native speaker. For this reason, L2 learners tend to have a "syntactic accent" (p. 57) in L2 sentence processing. Evidence for this "syntactic accent" is, for example, provided by studies showing that bilinguals oftentimes use L1 parsing strategies when processing their L2 (e.g., Liu, Bates, & Li, 1992; Sasaki, 1991) which will be discussed in more detail in section 3.2 on L1 transfer of (morpho)syntax. Further evidence for L1 transfer effects in different language domains will also be discussed in chapter 3.

However, MacWhinney (2005a) asserts that in the areas of morphosyntax, transfer rarely occurs as usually differences between languages are too big so that morphosyntactic features are not mapable and cannot be transferred (p. 55). He states that "[…] in morphosyntax, it is typically impossible to transfer from L1 to L2. For example, an English learner of German cannot use the English noun gender system as a basis for learning the German noun gender system. This is because English does not have a fully elaborated noun gender system." (p. 58). But, "On the other hand, there can be some real transfer effects to German from other languages that have full nominal gender systems. For example, a Spanish speaker might well want to refer to the moon as feminine on the basis of *la luna* in Spanish and produce the erroneous form *die Mond* in German rather than the correct masculine form *der Mond*." (p. 59). Hence, this means that transfer only occurs when it is possible. Important for the present thesis, the Competition Model explicitly predicts negative transfer effects when two languages differ regarding the gender of a noun, which is a prediction that will be tested in Experiment 1.

Nevertheless, the predictions of the Competition Model have rarely been tested on grammatical gender processing so that evidence in this area is scarce. Cue transfer and, as such, transfer of L1 processing strategies has mostly been investigated using the example of agent–patient identification depending on, for example, animacy cues or word order cues (see e.g., MacWhinney, 1997, p. 129; or Su, 2001, p. 85). Various studies with connectionist models have shown that acquisition of even complex nominal and declensional systems like the German system can successfully be modeled (Gupta & MacWhinney, 1992; Taraban, McDonald, & MacWhinney, 1989), that is, the models can actually learn the system based on cues derived from varying input.

The Competition Model emphasizes the role of cues and cue strength in language learning. Cue strength is a direct function of cue validity. Consequently, also in the learning of grammatical gender, the cue-based learning should be important. This would mean that the gender of transparent Spanish nouns should be especially easy to acquire because gender cues are both frequent and reliable. The gender of German nouns, on the other hand, should be more difficult to acquire because gender cues are either not available or not very reliable. The availability and frequency of a cue as well as its reliability is important. The characteristics of the German and Spanish gender system will be explained in more detail in section 4.1.1. In Experiment 1, similar and different L2 structures (i.e., in terms of gender congruency) are tested. So the Competition Model would predict that similar structures are easy, while different structures are difficult. Especially interesting in this regard is Experiment 2, as here a structure that is unique to L1 but not present in L2 is tested. It is not known what the predictions of the Competition Model are for this case.

The processing of transparent and intransparent Spanish nouns was investigated in the earlier cited study by Hernandez et al., (2004). They found increased activity for irregular nouns compared to regular nouns in early as well as late English–Spanish bilinguals, similar to previous monolingual studies. They also found increased activation in non-language areas in late bilinguals, which confirms the additional effort that has to be made to overcome L1 entrenchment. They interpret the differential effects found for regular and irregular nouns in terms of resonance and state that resonance is probably easier to achieve for regular than for irregular L2 items (p. 13). It is also remarkable that the English learners of Spanish were able to learn grammatical gender successfully, even though grammatical gender is a feature that is not present in their L1[19]. This is in accordance with the claim of the Competition Model that structures that are unique to the L2 are more easily acquired than structures that are different. In a later study, the greater activation for irregular items is interpreted in terms of less established resonance for irregular L2 items (Hernandez, Hofmann, & Kotz, 2007).

In conclusion, the DP Model as well as the Competition Model are attractive because they manage to provide a framework for L1 learning and processing and extend it to include L2 learning and processing. Both models provide an explanation for AoA effects, even if for different reasons. Nevertheless, the models could still be described as “optimistic”, that is, conceding that proficiency plays a role and can overcome age effects as discussed in section 1.4. Concerning the present research project, both models predict that L2 learners will have problems with grammatical gender, as a part of morphosyntax. The Competition Model further predicts transfer effects for both experiments of the present thesis. The more specific predictions will be described at a later point. A weakness of the DP Model is that it cannot account for L1 transfer effect. In the Competition Model, on the other hand, the exact reason for differential strength of entrenchment and discrepant success of L2 acquisition across language domains remains unclear. The next chapter will deal with grammatical gender and the difficulties L2 learners experience in this area of morphosyntax.

Chapter summary

Although in the case of L1 acquisition the CPH is less controversial, for L2 acquisition it is still a matter of debate whether there really is a CP or not. Yet, many researchers concur in that there are at least some effects of AoA even if they are probably less grave than postulated in the early days of L2 acquisition research. It is also evident that the effects are not equally strong for all linguistic domains. The acquisition of, for example, (morpho)syntax is more severely affected than the lexicon. Nonetheless, L2 speakers with a high proficiency are sometimes able to overcome the deficits caused by a higher AoA, even in the domain of (morpho)syntax. The DP Model and the Competition Model are each able to explain some of the AoA effects discussed in the present chapter and also some of the phenomena that will be discussed in the next chapters. In the following chapter, I am going to take a closer look at the category of grammatical gender, discussing research results trying to illuminate the reason for the pervasive difficulties with L2 gender.

[19] As will be discussed in section 3.3.1, native speakers of languages that lack gender are often at a disadvantage when learning an L2 with grammatical gender.

2. The difficulty of grammatical gender

As discussed in the previous chapter, the different domains of a language are differently affected by AoA effects. Current evidence suggests that, for example, the domain of grammar or (morpho)syntax is more severely impaired than the lexicon. However, also within the domain of (morpho)syntax several aspects that are differently affected by AoA can be distinguished. Clahsen and Felser (2006), for example, reviewing evidence from several experiments, arrive at the conclusion that brain response patterns for native speakers and high-proficient late bilinguals differ between some grammatical tasks – but not between all. They point out that it is not enough to declare that acquiring "grammar" is more difficult for late L2 learners than acquiring vocabulary because in some domains of grammar, adult L2 learners can very well achieve native-like proficiency (p. 568). Furthermore, it is necessary to find out which grammar domains are more and less difficult for L2 learners in order to gain a better understanding of the origin of difficulties in L2 acquisition. Based on the reviewed experimental evidence, they conclude that the processing of complex syntax (e.g. nonlocal dependencies, hierarchically complex structures) and morphosyntax does not become native-like in late L2 acquisition, even in very high-proficient learners. Furthermore, DeKeyser (2005), also reviewing evidence on late L2 grammar acquisition, draws the conclusion that in contrast to, for example, word order, L2 morphology (or morphosyntax), such as third person singular *-s* or plural formation remains hard for many adult L2 learners (p. 6). Hence, it is clear that the insight "grammar is difficult" is too broad and more carefully designed research needs to be conducted in order to find out what kind of structures are attainable for adult L2 learners and which structures tend to remain difficult even at native-like levels. As we shall see in the present chapter, one area of morphosyntax adult L2 learners frequently continue to struggle with is certainly grammatical gender and gender agreement (e.g. Corbett, 1991; Lemhöfer, Schriefers, & Hanique, 2010, p. 150). I will review findings on the difficulty of learning grammatical gender in an L2 and discuss factors that influence L2 gender processing.

2.1 L2 gender is difficult for late learners

One example for the difficulty of L2 grammatical gender is provided in the aforementioned study by Davidson and Indefrey (2009, cf. section 1.4.2). They presented results of Dutch learners who after short exposure acquired sensitivity to declension violations in German but not to grammatical gender violations. The authors stated that "[it] is possible that grammatical gender is more difficult to acquire than other grammatical distinctions" (p. 444). Similarly, Dewaele and Véronique (2001) conducted a study on gender errors in L2 French with 27 native speakers of Dutch. Subjects were advanced speakers of French. They had learned French as an L2 or L3 in secondary school for four to six years and had been currently enrolled in an intensive French course for the last five months. The researchers conducted and recorded five hours of interviews in total, with 17,613 words and 9,378 modifiers. In total, 516 gender errors were made[1]. Dewaele and Véronique (2001) concluded that "gender errors are abundant in L2 production"[2] (p. 275). This result is confirmed by, for example, Franceschina (2001). She reported the striking case of a native English speaker who had lived in

[1] The authors state that since (in oral French) the masculine and feminine form is the same for some adjectives, the actual amount of gender errors might have been even higher (p. 283).

[2] Note that the amount of gender errors was not compared to any other type of error so it is difficult to estimate whether gender errors were committed more often than other types of errors.

Spanish-speaking countries for a total of 24 years and continued to have problems with Spanish grammatical gender agreement but less so with number agreement. Agreement errors in adjectives, articles, pronouns, and demonstratives in 94 minutes of recorded natural conversation were investigated. Only 7 % of agreement errors were number agreement errors, while 93 % were gender agreement errors. Consequently, Franceschina (2001) claims that "Gender agreement is clearly more problematic than number in every case" (p. 236). These difficulties of L2 speakers are remarkable as L1 speakers rarely make gender errors (e.g. Caselli, Leonard, Volterra & Campagnoli, 1993). Furthermore, research has shown that L1 speakers are able to efficiently use gender information as valuable cues in online-processing (Bates, Devescovi, Hernandez, & Pizzamiglio, 1996; Dahan, Swingley, Tanenhaus, & Magnuson, 2000; Guillelmon & Grosjean, 2001; Lew-Williams & Fernald, 2010; Scherag, Demuth, Rösler, Neville, & Röder, 2004; Wicha, Moreno, & Kutas, 2004). This is even true for young children, as demonstrated in a study by Lew-Williams and Fernald (2007). They showed that native Spanish-speaking children as young as three years old were able to use L1 gender information on determiners to identify objects in a "looking-while-listening" procedure (Fernald, Perfors, & Marchman, 2006; this procedure will be explained in more detail below).

In contrast, late L2 speakers´ performance is not only impaired in overt L2 production, but also in the processing of gender violations in L2 comprehension. For example, Scherag, Demuth, Rösler, Neville, and Röder (2004) investigated possibly differential effects on semantics and morphosyntax (gender agreement) in late L2 acquisition (next to effects of language attrition which will not be discussed here). Native English long-term immigrants to Germany performed an auditory LDT with semantic and morphosyntactic adjective primes. Semantic adjective primes were either semantically related or unrelated to the target noun, morphosyntactic adjective primes either agreed with the target noun´s gender or not. Results showed that native German speakers benefitted from semantic as well as morphosyntactic adjective primes, while native English speakers were only sensitive to semantic but not to morphosyntactic adjective primes. So it seems that gender agreement of adjectives as part of morphosyntax is difficult to learn for late L2 speakers. Guillelmon and Grosjean (2001) arrived at a similar conclusion using a comparable method with an auditory naming task. After listening to short phrases with the following structure: gender-congruent/gender-incongruent/neutral baseline determiner–adjective–noun, subjects had to repeat the noun as quickly as possible. Early bilinguals and monolinguals showed significantly faster naming times when the determiner was congruent and significantly slower naming times when the determiner was incongruent, compared to the baseline. However, high-proficient late English–French bilinguals´ recognition of a noun was not affected by the gender congruency of the preceding determiner. Consequently, similar to the previous study, late bilinguals were insensitive to the gender agreement of a determiner preceding a noun.

These results were confirmed in an eye-tracking study by Lew-Williams and Fernald (2010). They used no gender violations but investigated whether L2 speakers could benefit from L1 gender cues in aural comprehension in the same way as L1 speakers. Subjects were native speakers of English who were of advanced proficiency in their L2 Spanish and a native Spanish-speaking control group. The "looking-while-listening" procedure also used in the aforementioned study investigating gender processing in native Spanish-speaking toddlers by Lew-Williams and Fernald (2007) was employed. In response to the Spanish stimulus sentences, for example: *Encuentra la pelota!* (*Find the$_{\text{-}fem}$ ball$_{\text{-}fem}$*!), subjects had to look at the correct picture of the two pictures presented (experiment 1). Pictures either had the same gender or different genders. Just as the three-year-olds in the study of Lew-Williams and Fernald (2007), and even more efficiently, native speakers in the study of Lew-Williams and Fernald (2010) were able to identify the correct picture faster at different-gender trials than at same-gender trials. This is due to the fact that in these cases the gender of the determiner provided an informative cue for identifying the critical picture. L2 speakers, however, showed no differences in reaction times (RTs) and were not able to benefit from the gender information provided by the determiner.

Sensitivity to L2 gender agreement has also been investigated using ERPs. For example, in a previously cited study (cf. section 1.5.2), Tokowicz and MacWhinney (2005) investigated processing of different structures, including determiner–noun gender agreement. Subjects were native English speakers and low-proficient in their L2 Spanish. They had learned between one and four semesters of Spanish in college. Items included violations of tense-marking/auxiliary omission which is similar in L1 and L2, determiner number agreement, which is different in the two languages and determiner gender agreement, which only exists in L2. They found a native-like P600 in response to violations of similar and unique L2 structures, that is, tense-marking and gender agreement violations but not in response to determiner–number agreement. This contradicts the usual finding that L1 speakers of English have more problems with gender than with number agreement (Franceschina, 2001; Gillon Dowens, Vergara, Barber, & Carreiras, 2010; Sagarra & Herschensohn, 2010 (only GJT); White, Valenzuela, Kozlowska–Macgregor, & Leung, 2004). The authors interpret these findings with regard to L1 influences in L2 processing and L1 and L2 similarity which will be discussed in more detail in chapter 3. Nevertheless, more important at this point is the fact that determiner–noun gender agreement was only native-like in the online measure but not in the offline GJT. Accuracy rates in the GJT were at chance for all structures and significantly lower for gender agreement, which is in line with the results of the GJT by Sagarra and Herschensohn (2010).

Some more ERP studies researching gender processing were conducted by Foucart and Frenck-Mestre (2011, 2012). They investigated processing of gender agreement violations (determiner–noun and adjective–noun agreement) of high-proficient L1 German (2011) and L1 English (2012) speakers in L2 French. In the 2011 study, they tested determiner–noun agreement as well as (pre-posed vs. post-posed) attributive adjective–noun agreement. In the 2012 study, they tested (pre-posed vs. post-posed) attributive and predicative adjective–noun agreement. L2 speakers only processed some of the violations, the determiner–noun (2011) and the post-posed adjective noun (2012) structures in a native-like way, as evidenced by a P600 also found in the native control groups. For the other agreement structures, however, L2 speakers did not show a P600 in response to violations whereas the control group did. Likewise, Gillon Dowens, Guo, Guo, Barber, and Carreiras, (2011) and Gillon Dowens, Vergara, Barber, and Carreiras (2010) found significant differences in online processing of gender agreement violations in determiner–noun and noun–adjective constructions in high-proficient late L2 speakers. They investigated gender and number processing of L1 English (Gillon Dowens et al., 2010) and L1 Chinese (Gillon-Dowens et al., 2011) subjects in L2 Spanish compared to a native Spanish control group (Gillon-Dowens et al., 2010). Also in these studies, L2 speakers differed from native speakers. English speakers processed determiner–noun constructions in a native-like way as evidenced by a LAN–P600 pattern but only exhibited a P600 and no LAN in response to noun–adjective violations. The native Chinese group (Gillon-Dowens et al., 2011), on the other hand, showed only a P600 and no LAN to both constructions. Furthermore, the native English group (Gillon-Dowens et al., 2010) exhibited greater difficulties with gender than number processing, as evidenced by differences in both components´ amplitude, distribution, and latency as well as higher error rates for gender agreement in sentence acceptability judgments. Thus, all of the studies presented in this paragraph and the previous paragraphs provide evidence for differences in gender agreement processing between native and late L2 speakers. However, in some cases, L2 processing was native-like. Some of the studies showed that sensitivity to gender violations was also influenced by proficiency effects (Foucart & Frenck-Mestre, 2011), L1 transfer effects (Foucart & Frenck-Mestre, 2011; Gillon Dowens et al., 2010; Tokowicz & MacWhinney, 2005) and effects of agreement distance (Foucart & Frenck-Mestre, 2012). Consequently, AoA does not seem to be the only factor affecting L2 gender processing. In the next section, I will take a closer look at some of the other factors that play a role in L2 gender processing. L1 transfer in grammatical gender processing will be discussed in detail in section 3.3.

2.2 Factors affecting L2 (gender) processing

In the following, I will discuss some of the factors which have been shown to play a role in L2 gender processing, L2 processing in general, and sometimes maybe even in L1 processing. I will look at the influence of agreement distance, type of task or task demands, and proficiency effects. It will become clear that some of these aspects are interrelated and can sometimes be difficult to separate. Other factors of importance are L1 influences and L2 characteristics which will, however, only be briefly considered here as chapter 3 specifically deals with L1 transfer effects. Furthermore, even if the studies presented in this section investigate more specific aspects of L2 gender (agreement) processing, they nevertheless also provide additional evidence for differences between native and late L2 gender processing in general because they usually include a native speaker control group.

Agreement distance

One of the factors affecting L2 gender processing which has been heavily investigated is agreement distance. Agreement distance refers to the structural distance across which agreement between two (or more) elements has to be established. The agreeing elements can be relatively close to each other, such as a noun and its determiner, or structurally more distant from each other, such as a noun and its relative pronoun. As mentioned at the beginning of this chapter, the importance of agreement distance in L2 processing has already been underlined in the review article of Clahsen and Felser (2006). They claimed that non-local dependencies remain difficult even for native-like L2 speakers. Furthermore, agreement distance has also been shown to affect L1 processing (Barber & Carreiras, 2005). As mentioned in section 1.5.1, it is possible that the processing of more distant agreeing elements requires syntactic computation and thus reliance on the procedural memory system which according to the DP Model is supposed to be impaired in late L2 learning. The gender of bare nouns or mere determiner agreement, on the other hand, might also be quite successful using lexical (Blom et al., 2008; Kempe et al., 2010; Sabourin et al., 2006) and thus, declarative strategies which are supposed to be not impaired in late learners.

In the following, I will discuss a few studies that provide evidence for greater processing problems with increasing agreement distance in L2 speakers. The agreement distances investigated in the different studies vary. Some studies compared agreement processing within the NP, that is, determiner–noun or noun–attributive adjective agreement. Other studies investigated differences between NP and verb phrase processing, while again others also investigated processing of even more distant relationships, such as relative clauses and their antecedents. As always, different methods, ranging from offline paper-and-pencil questionnaires to behavioral studies to eye-tracking and ERP studies investigating online processing, were employed.

Myles (1995), for example, found effects of agreement distance in an oral repetition task (experiment 1) and an error correction task (experiment 2). Level of embeddedness was manipulated in questions with interrogatives (experiment 1) and in noun–adjective agreement (experiment 2).

Subjects were native English and L2 speakers of French with different proficiencies. In experiment 1, subjects were advanced speakers of French who had learned French for seven years in school (no stay abroad mentioned) and had to repeat the auditorily presented questions and answer them. In experiment 2, subjects had learned French for three to eight years and had to correct a written text containing different types of errors. The critical errors were gender agreement errors with varying structural distances: within the NP, within the clause, outside the clause (relative pronouns), and outside the verb and NP (adjectival sentence in apposition). For both experiments it was found that structural agreement distance predicted accuracy, that is, the greater agreement distance was, the more errors were made (experiment 1) and overlooked (experiment 2). Experiment 2 also showed that this effect was potentially mediated by proficiency. A working memory explanation was put forward, arguing that with increasing proficiency and hence increasing automaticity more working memory resources for correctly referencing more distant elements become available.

Using different language pairs, also Sabourin, Stowe, and De Haan (2006) provided evidence for effects of agreement distance in an offline task. Their L2 learners of Dutch had greater difficulties with judging the correctness of gender agreements in relative clauses (experiment 2) than with a simple gender assignment task (experiment 1). Both experimental tasks were presented offline, as pen-and-paper questionnaires and subjects had different language backgrounds (German, Romance, and English). In the first experiment, total accuracy at assigning the correct article to nouns was above 80 %. In the second experiment, accuracy dropped by at least 10 % for each language group. Sabourin et al. (2006) also found effects of language background, which will be discussed more thoroughly in the next chapter, and item familiarity. Interestingly, the Dutch control group did not exhibit perfect accuracy in the second experiment, either. They scored only at around 98 %[3] on average which is surprising considering that time pressure was low in this offline task. This shows that even if native speakers perform significantly better than L2 speakers, they are not necessarily perfect when establishing gender agreement across clausal boundaries.

These results are similar to Keating´s (2009), who carried out an eye-tracking study with native English speakers of different proficiencies in their L2 Spanish and a native control group. He investigated processing of gender agreement in three different sentence contexts: within the determiner phrase, within the verb phrase and across the phrase boundary in a subordinate phrase. L2 subjects were beginning, intermediate, and advanced subjects. The advanced subjects were very high-proficient, had lived or studied in a Spanish-speaking country, held a college degree in Spanish, and/or pursued an academic career in Spanish so that self-rated proficiency as well as daily exposure to Spanish was high. The results showed that advanced learners of Spanish were sensitive to gender agreement violations on Spanish adjectives within the determiner phrase, just as native speakers. Therefore, Keating (200) concluded that gender agreement is acquirable for late L2 learners but is probably acquired late as beginning and intermediate learners were not sensitive to agreement violations within the determiner phrase. However, when the structural distance between nouns and their modifying adjectives was increased beyond the NP, L2 speakers as a group differed from native speakers even though there were some cases of native-like processing. Keating assumes that differences in working memory capacity could play a role in learners´ sensitivity to non-local agreement errors.

[3] As overall accuracy values are only presented in bar diagrams and not as absolute values, this value is an estimate taken from Figure 2 (Sabourin, Stowe, & de Haan, 2006, p. 18).

Effects of L2 gender agreement distance have also been investigated using ERPs. The earlier cited study by Foucart and Frenck-Mestre (2012) investigated processing of noun–adjective (attribute vs. predicative) gender agreement violations. Subjects were high-proficient L1 English–L2 French speakers studying at a French university and were compared to a native control group. Foucart and Frenck-Mestre (2012) showed that violations between the noun and a post-posed adjective elicited a P600 for the native as well as the late L2 group (more frontal for L2). Violations between the noun and a pre-posed adjective (a less frequent structure in French), on the other hand, exhibited a P600 in natives but only an N400 in the L2 group. In response to violations between the noun and a predicative adjective, natives presented a P600, while no effect was found for the L2 group. Apparently, processing non-local structures is more difficult for L2 speakers. Nonetheless, in an eye-tracking version of this task results were similar for the two groups. This difference in results strongly supports the explanation of a potential working memory overload presented in the earlier mentioned studies by Keating (2009) and Myles (1995) since the ERP study naturally was not self-paced, whereas in the eye-tracking studies reading speed and re-reading could be adapted to individual processing needs. The evidence concerning the difference in results for the post-posed and pre-posed adjectives, however, remains inconclusive. As Foucart and Frenck-Mestre state, it is possible that post-posed adjectives were easier for L2 learners because they are more frequent in French. Therefore, in classroom instruction more attention is apparently paid to this structure and post-posed adjectives are acquired before pre-posed adjectives, reinforcing their processing advantage. It is also possible that for pre-posed adjectives, a structure also present in English, L1 influences were stronger than for post-posed adjectives, which is a structure that is not present in English. In any case, next to the effects of agreement distance, once more it was shown that gender (agreement) processing for high-proficient late L2 speakers is not native-like.

So even if a great deal of studies presents evidence in favor of greater difficulties for late L2 learners with increasing structural distance, effects of agreement distance are not always found. In the aforementioned study by Dewaele and Véronique (2001) investigating oral gender errors of native Dutch speakers in their L2 French, no significant differences between structural distance (attributive adjectives in anteposition and postposition vs. predicative adjectives) were found. Even though error rates for predicative adjectives were higher, this difference proved to be non-significant, possibly due to the high standard deviation. However, a significant difference in accuracy between determiner–noun and noun–adjective agreement was found. These authors concluded that agreement distance is irrelevant for advanced speakers who have mastered a certain L2 structure. It has to be noted, though, that contrary to the previously discussed studies, they did not specifically manipulate agreement distance but rather obtained their data from unstructured interviews. In another corpus analysis, Bartning (2000) analyzed oral production data of Swedish learners of L2 French. She also failed to find a difference in accuracy between attributive and predicative adjective agreement. However, oral production data of open interviews without intentional manipulation of agreement distance might not be a good source of information on difficulties with agreement distance: Research has shown that L2 learners tend to reduce their error rates by avoiding structures they are not sure how to use (Hubert, 2011; Kleinmann, 1977).

Despite these findings, many experimental online as well as offline studies carefully manipulating agreement distance have presented evidence for effects of agreement distance. These effects are probably caused by the greater working memory load imposed by establishing reference across more distant structure (Foucart & Frenck-Mestre, 2012; Keating, 2009; Myles, 1995) and mediated by proficiency. Even if native speakers have also been shown to have greater problems with greater structural distances, late L2 speakers´ performance in the processing of gender agreement has only been shown to be native-like in very simple tasks, that is, local and canonical structures (Foucart & Frenck-Mestre, 2012; Gillon Dowens et al., 2010; Keating, 2009; only native-like in ERPs but not in the GJT: Tokowicz & MacWhinney, 2005). In the present thesis, agreement distance is varied between the two experiments. In the first experiment, processing of gender agreement within the NP (determiner–noun agreement) is examined whereas the second experiment investigates gender agreement across sentences boundaries, namely, in pronoun resolution. However, it has become clear that agreement distance is not the only factor that influences processing of gender agreement besides AoA. Some of the studies discussed have also found influences of L1 (Foucart & Frenck-Mestre, 2012; Gillon Dowens et al., 2010; Sabourin et al., 2006). Since L1 influences and the interplay of L1 and L2 characteristics regarding L2 gender processing are the central topic of the present thesis, L1 transfer effects will be discussed in more detail in chapter 3. The next section will deal with another factor that influences L2 gender processing, namely differing task demands in comprehension vs. production and online vs. offline tasks. As we shall see, working memory is also thought to play a role here.

Task demands

Montrul, Foote, and Perpiñán (2008) examined differences between reading comprehension and oral production in gender agreement processing of noun-drop structures. In three experiments, performance in reading comprehension, which is offline and untimed, was compared to oral production, which is online and naturally time-constrained. To this end, a sentence–picture matching task (reading comprehension) developed by White, Valenzuela, Kozlowska–Macgregor, and Leung (2004), a gender recognition task (reading comprehension) and an oral picture description task were employed. Subjects were native English speakers with Spanish as an L2, native Spanish speakers as a control group, and so-called “heritage speakers”, referring to Mexicans born and raised and schooled in the US. The L1 of those “heritage speakers” was Spanish which was also spoken at home, but their first exposure to English was before the age of 5 in pre-school. Their results showed that both bilingual groups, the native English speakers as well as the “heritage speakers”, made gender agreement errors. However, L2 learners made more errors in the oral production task, while the L2 speakers of Spanish made more errors in the reading comprehension tasks. This can be explained by the fact that heritage speakers have more oral than written practice in Spanish because they never received formal instruction in Spanish. Nevertheless, more important at this point is that the present study shows that the L2 speakers had more problems with grammatical gender in the oral production task than in the reading comprehension tasks. In the offline comprehension tasks, L2 speakers had accuracy scores above 80 % and some even performed within the range of native speakers. Yet, in the earlier cited study by White et al. (2004)also investigating performance in L2 Spanish, L2 subjects performed equally well in the online comprehension task (sentence–picture matching task) and an oral production task (a picture description task). Montrul et al. (2008) claim that this is could be due to a ceiling effect because White et al.´s (2004) picture description task was easier, as only high-frequent nouns with transparent gender-endings were used. Hence, it can be concluded that L2 gender agreement is more difficult in oral production than in offline comprehension tasks.

However, as Grüter, Lew-Williams, and Fernald (2012) argued, studies comparing oral production and offline reading comprehension contain an important confound: As production studies are usually conducted online, it is difficult to decide whether these difficulties arise from its online nature or from problems specific to production. In order to disentangle these factors and to dispose of the time pressure confound, Grüter et al. investigated differences in grammatical gender processing between production and comprehension as well as online and offline tasks. Subjects were native English speakers who were high-proficient in their L2 Spanish and a control group of native Spanish speakers. The authors found that in the offline comprehension task, a sentence–picture matching task (Montrul et al., 2008, experiment 1; White et al., 2004), both groups performed at ceiling, replicating the findings by Montrul et al. and White et al. There were no differences in accuracy between the two groups. In the second experiment, however, an elicited production task, the native group had a mean accuracy of 98.7 %, while the L2 group only obtained an accuracy of 80 %, which was significantly worse than the native group. Experiment 3 was a "looking-while-listening" procedure, an online comprehension eye-tracking task also used by Lew-Williams and Fernald (2007, 2010). In response to a Spanish stimulus sentence, for example, *Encuentra la pelota*! (*Find the ball!*), subjects had to look at the correct picture of the two pictures presented. The depicted objects either had the same grammatical gender or different genders. Similar to the results of the L1 and L2 subjects of Lew-Williams and Fernald (2010) and the monolingual toddlers of Lew-Williams and Fernald (2007), both participant groups were able to use gender cues to speed up processing, reacting faster to different-gender than same-gender trials. There was no interaction effect of trial type and language group, but planned comparisons within language groups revealed that the difference between trial types was only significant for the L1 group and not for the L2 group. This suggests that native speakers were able to use gender cues more efficiently than L2 speakers. Overall, problems with gender in L2 processing became more apparent in a production and online comprehension task than in offline comprehension tasks. Moreover, online comprehension proved to be more difficult than offline comprehension. Grüter et al. (2012) concluded that L2 learner´s difficulty with gender in production tasks is not so much a problem caused by characteristics inherent to production but probably rather due to time constraints inherent to online tasks, rendering correct gender retrieval difficult.

Thus, L2 gender processing is affected by the different task demands that arise in comprehension and production, offline and online tasks. Thereby, the difference in time pressure between online and offline tasks and production and comprehension tasks appears to be more important than the modality tested. Apparently, production tasks are more difficult than comprehension tasks and online tasks more difficult than offline tasks, due to time pressure. Native-like performance of L2 speakers is often found in offline comprehension tasks (Grüter et al., 2012; Montrul et al., 2008). Therefore, when doing research on gender processing, it is important to take into account the possible influences of task demands. In the present thesis, different tasks are employed across the two experiments. In the first experiment, an offline task, as well as two online tasks (comprehension and production), are employed. The tasks are an offline gender assignment task without time pressure measuring error rates, a lexical decision task (LDT) and a picture naming task (PNT) measuring error rates as well as RTs. In the second experiment, an online reading task including a GJT and measuring ERPs as well as accuracy rates[4] is used. However, next to agreement distance and task demands, another factor has been shown to play a role in gender agreement processing and in L2 processing in general (cf. section 1.4), namely, proficiency. Some of the studies discussed earlier found evidence for proficiency effects (Keating, 2009; Myles, 1995, experiment 2) and in the next section proficiency effects in L2 gender processing will be discussed.

[4] RTs were not analyzed for reasons which are explained in more detail in section 5.2.4.

Proficiency

In the previously cited eye-tracking study, Keating (2009) found that advanced learners of L2 Spanish with English as their L1 were sensitive to gender agreement violations of Spanish adjectives within the NP, just as native speakers. Beginning and intermediate subjects, on the other hand, were insensitive to gender violations. When agreement violations occurred outside the determiner phrase, that is, within the verb phrase and in a subordinate clause, all L2 proficiency groups performed worse than native speakers. Nevertheless, there were also some cases of native-like processing outside the determiner phrase. Myles (1995) conducted an offline error correction task as his second experiment and also found that error rates with increasing structural distance correlated negatively with proficiency. In other words, the lower the proficiency, the more errors were overlooked with increasing structural distance. Furthermore, in a self-paced reading task, Sagarra and Herschensohn (2010) found that beginning learners showed no sensitivity to gender and number violations, whereas the intermediate group processed these violations similar to the native group.

Hence, just as in other grammatical tasks, as discussed in chapter 1, effects of proficiency are also found in L2 gender processing. Thus, even if gender processing is difficult, it clearly does improve as proficiency increases. Native-like processing, however, is rarely found for gender processing. As mentioned before, it seems that native-like processing can be found only for local and canonical structures, for example, agreement within the determiner phrase (Foucart & Frenck-Mestre, 2012; Gillon Dowens et al., 2010; Keating, 2009; only native-like in ERPs but not in the GJT: Tokowicz & MacWhinney, 2005) or if task demands are low as, for example, in offline reading comprehension (Grüter et al., 2012; Montrul et al., 2008).

Chapter summary

Gender is an aspect of grammar that is difficult to acquire in an L2. Research has shown that L2 gender processing is rarely native-like and more vulnerable than native processing. L2 performance depends on a trade-off of many factors which put different strains on working memory, such as agreement distance and task demands (e.g. online vs. offline processing, production vs. comprehension, general time pressure). Just as we saw in chapter 1 discussing (morpho)syntactic processing in general, proficiency also plays an important role in L2 gender processing. Even if grammatical gender is difficult to acquire for L2 learners, instances of native-like gender processing in high-proficient late L2 speakers can be observed when structural distance and time pressure are low, such as determiner–noun agreement in offline comprehension tasks. Yet, with increasing agreement distance and time pressure, such as in online production tasks, differences between L1 and L2 speakers become more and more apparent. As mentioned before, the present thesis manipulates all of these aspects which have been shown to play a role in L2 gender processing across two experiments. Gender processing is investigated in offline and online tasks, production and comprehension tasks, in determiner–noun agreement, and agreement across sentence boundaries. Furthermore, L1 influences have also been shown to play a part in L2 gender processing and will be further discussed in the next chapter.

3. L1 transfer effects in L2 lexicon, (morpho)syntax, and grammatical gender

As we have seen in the previous chapters, it is difficult for late L2 learners to become native-like and this seems to be truer for some aspects of a language than for others. Domains that are especially affected are (morpho)syntax and in particular grammatical gender. Explanations for AoA effects range from CP accounts and postulations of maturational effects to emergentist accounts, which hold L1 transfer effects responsible due to L1 entrenchment (cf. section 1.5.2). In their review article on ultimate attainment in L2 processing, Clahsen and Felser (2006a) discuss four main explanatory attempts for L1 and L2 processing differences: lack of relevant grammatical knowledge (i.e., lack of competence), influence from the L1, cognitive resource limitations, and maturational changes after puberty (p. 564). The present dissertation focuses on the second explanatory attempt: L1 influences during L2 processing with regard to grammatical gender. As we shall see, there is abundant evidence for L1 transfer impairing L2 processing. In the literature, L1 transfer effects have been reported for the domains of phonology, the lexicon, and morphosyntax (Flege, Bohn, & Jang, 1997; Odlin, 1989). In the present chapter, I will briefly discuss results of studies investigating transfer effects in the bilingual lexicon, then give a short overview over studies investigating transfer effects in the domain of syntax and morphosyntax/morphology, and finally present a more extensive discussion of transfer effects in grammatical gender, which is the center of interest of the present thesis.

First of all, what does language *transfer* actually mean? According to the (working) definition given by Odlin (1989), transfer is "[…] the influence resulting from similarities and differences between the target language and any other language that has been previously (and perhaps imperfectly) acquired." (p. 27). In the present chapter and in the present thesis, mainly transfer stemming from the L1 will be considered. Another widely used term is language *interference*. A definition of interference is also provided by Odlin (1989): "The term interference implies no more than what another term, negative transfer, does, but there is an advantage in using the latter term since it can be contrasted with positive transfer, which is the facilitating influence of cognate vocabulary or any other similarities between the native and target languages." (p. 26). So importantly, L1 transfer does not necessarily have only negative consequences and does not always lead to erroneous behavior. In the present thesis, the terms *negative transfer* and *interference* will be used almost synonymously; however, I will reserve the term *interference* for online effects. As discussed in section 1.5.2, according to the Competition Model, transfer occurs because of L1 activation and competition during L2 processing (Hernandez et al., 2005; MacWhinney, 2005a).

3.1 Lexical competition and transfer

There is ample and robust evidence, involving a wide variety of methodologies and language pairs, for the parallel activation of the two languages of a bilingual and for interference at the lexicon level[1]. If L1 words, or more specifically, L1 nouns, are activated when processing L2 nouns, this might also have consequences for the processing of grammatical gender in L2. According to some speech-processing models, grammatical gender is a lexical-syntactic property of nouns stored at the lemma-level (Levelt, Roelofs, & Meyer, 1999; Levelt, 1989; Schriefers & Jescheniak, 1999). If upon retrieval of

[1] For a more in-depth discussion of cross-language activation in the bilingual lexicon see, for example, Dijkstra (2005) and Altarriba and Gianico (2003).

an L2 noun its L1 translation equivalent is also retrieved, it seems easy to imagine that the L1 gender information will also be activated. For this reason, a few examples of cross-lexical activation between languages will be given here. As we shall see, just as in transfer studies in other areas, proficiency effects are also found here.

A good example for the activation of L1 words influencing L2 (reading) processing in high-proficient late bilinguals is provided by Elston-Güttler and Williams (2008). In an anomaly detection task with English sentences including German homonyms that have two different English translations (e.g. *Blase* meaning *blister* or *bubble*), they showed that L1 polysemy can affect L2 meaning interpretation. This indicates that L1 words are activated during L2 processing and that the lexica of the two languages of a bilingual might not be completely separate in the bilingual brain[2]. These results are in line with the results of two eye-tracking experiments involving auditory sentence processing in Russian–English late bilinguals by Marian and Spivey (2003a, cf. also 2003b). They found evidence for within-language competition as well as between-language competition. Further evidence for competition through L1 (Dutch) activation in L2 (English) auditory processing is provided in another eye-tracking study by Weber and Cutler (2004). In addition, it is important to point out that in the study by Marian and Spivey (2003a), cross-language competition also occurred when languages were as dissimilar as Russian and English. Apparently, non-target language activation even arises when languages differ in script as, for example, with English and Chinese or in modality as, for example, with spoken language and American Sign Language (Kroll & Bogulski, 2013, p. 4).

Traces of L1 influence in L2 noun processing are also found in ERP studies. Elston-Güttler, Paulmann, and Kotz (2005), for example, conducted a semantically primed LDT, measuring ERPs and RTs. Primes were translation equivalents of German homonyms[3] (e.g. *Kiefer* meaning *pine* or *jaw* in English). They found interference effects for L2 words presented without context and in a sentence context for low-proficient learners, but only for words presented without context for high-proficient learners. Interpreted within the Competition Model framework, this could mean that high-proficient subjects were better able to make use of L2 resonance provided by the sentence context in order to overcome L1 interference. This shows once more the importance of proficiency in L2 processing, also regarding L1 activation and interference, which is confirmed in a behavioral study by Degani, Prior, and Tokowicz (2011). They investigated the influence of Hebrew homonyms on processing of their L2 English translation equivalents in high-proficient late bilinguals and found L1 transfer effects. Interestingly, this effect was also found for late bilinguals whose L1 was English and who were high-proficient in their L2 Hebrew. Apparently, semantic transfer does not only occur from L1 to L2 but can also occur in the other direction in high-proficient bilinguals. Hence, the role of proficiency in language transfer cannot be overestimated. Further evidence for the L2 exerting influence on L1 processing is provided by van Hell and Dijkstra (2002). Nevertheless, according to Dijkstra (2005), L1 influences seem to continue even at higher proficiency levels.

All these results show that, to quote Grosjean (1989), “the bilingual is not two monolinguals in one person” (p. 3), but that the two languages of a bilingual or even the multiple languages of a

[2] For further discussion of language-nonselective access models for bilingual word processing see, for example, the Revised Hierarchical Model (Kroll & Stewart, 1994; Kroll & Tokowicz, 2005) or the BIA Model (van Heuven, Dijkstra, & Grainger, 1998) which has been extended to the BIA+ Model (Dijkstra & van Heuven, 2002; Dijkstra, 2005), for a review see Kroll and Bogulski (2013) and Kroll and Dussias (2006).

[3] Since no further explanations or examples of the stimuli are given by the authors, it is not clear whether the difference between polysemy and homonymy was strictly controlled in the two studies discussed here. Therefore, I will adhere to the terms used by the authors.

multilingual are not completely separate but influence each other[4]. Or as Kroll and Bogulski (2013) put it "[…] the available evidence suggests that, even when bilinguals intend to speak one of their two languages, information about lexical candidates in the other language is activated […]. Quite counter-intuitively, bilinguals do not appear to be able to restrict activation to words in the language they intend to speak" (p. 4). These effects occur even when languages are dissimilar and are mediated by proficiency. As stated before, if L1 transfer effects occur at the word level, it is conceivable that also grammatical gender information will be transferred. The transfer of grammatical gender information will be discussed in section 3.3. Next, I will look at transfer effects in other areas of (morpho)syntax.

3.2 L1 transfer of (morpho)syntax

After research on the bilingual lexicon, the interest in L1 (morpho)syntactic transfer has grown. While lexical transfer is intuitively easy to imagine, regarding (morpho)syntax, MacWhinney (2005) states that "There is clear evidence for massive transfer in audition, articulation, lexicon, sentence interpretation, and pragmatics. In the area of morphosyntax and sentence production, transfer is not as massive, largely because it is more difficult to construct the relations between L1 and L2 forms in these areas." (p. 55). So it seems that in the area of (morpho)syntax, it is more difficult to observe L1 transfer. For this reason, the present thesis aims to provide more insights into L1 transfer effects in an area of morphosyntax, namely, grammatical gender. Nevertheless, also in the case of morphosyntax and syntax, shared representations for L1 and L2 have been proposed by researchers. According to the "shared syntax" account, bilinguals have a single syntactic representation for structures that exist in both languages, whereas "separate syntax" accounts claim that the representations for these structures are language-specific (Hartsuiker, Pickering, & Veltkamp, 2004, p. 409). In the following, I will give an overview of some of the existing evidence. As we shall see, there is also evidence for L1 syntactic and morphosyntactic influences on L2 sentence processing even though the evidence is less clear-cut than in the case of the lexicon. Morphosyntactic transfer in a sentence context is especially important for the second experiment which investigates gender transfer in sentence processing. The present section also provides a theoretical background on (morpho)syntactic transfer in general, against which grammatical gender transfer studies discussed in the next sections can be evaluated.

3.2.1 L1 transfer of syntax

Evidence against a strict separation of L1 and L2 grammatical representations and for L1 influences also in grammatical processes comes from syntactic priming. The notion of a "shared-syntax account" or a "separate-syntax account" (Hartsuiker, Pickering, & Veltkamp, 2004) has been investigated across modalities with a wide variety of priming methodologies and grammatical constructions, such as active/passive constructions (Hartsuiker et al., 2004; Kantola & van Gompel, 2011; Weber & Indefrey, 2009), dative constructions (Loebell & Bock, 2003; Salamoura & Williams, 2007a; Shin & Christianson, 2009) and ambiguous relative clause attachment (Desmet & Declercq, 2006; Nitschke, Kidd, & Serratrice, 2010). The L2 at test was mostly English[5], but syntactic priming has been observed with numerous L1s (e.g. Spanish, German, Swedish, Greek, Korean, Dutch). This shows that not only L1 words, but also L1 syntactic information is active during L2 processing and influences L2 sentence

[4] This seems to be especially true for cognates as cross-language influences have been found to be even stronger for cognates than for noncognates (e.g., Costa, Caramazza, & Sebastián-Gallés, 2000; Dijkstra, Grainger, & van Heuven, 1999).

[5] Except for Nitschke et al. (2010) who used English as L1 and German and Italian as L2s.

processing, which has important consequences for the second experiment of the present thesis (chapter 4.2). Under a cognitive psychological account, ambiguous relative clause attachment might be especially relevant for my second experiment because, just as in anaphor resolution, a referent has to be stored in short-term memory and the correct referent has to be retrieved in order to understand the ensuing phrase or sentence. It is interesting that L1 information or processing preferences can possibly interfere during the retrieval of a referent stored in working memory.

Ambiguous relative clause resolution by L2 speakers has also been investigated using other methods such as eye-tracking or self-paced reading studies, which allow a more direct observation of readers´ parsing strategies. An often stated example of an ambiguous relative clause is the sentence *Someone shot the servant of the actress who was on the balcony,* which was first used by Cuetos and Mitchell (1988). In this sentence, it is unclear if the servant or the actress was on the balcony and languages differ with regard to their attachment preferences. The evidence regarding L1 transfer is more mixed than in the previously discussed studies. Frenck-Mestre and Pynte (1997)[6] and Juffs (1998) found evidence for transfer of L1 parsing strategies, while Felser, Roberts, Marinis, and Gross (2003) and Papadopoulou and Clahsen (2003) found no evidence for L1 transfer effects. Dussias and Sagarra (2007), on the other hand, even found evidence for transfer of L2 parsing strategies into L1 in high-proficient subjects. This is consistent with the suggestion by Nitschke et al. (2010) that the lack of finding transfer effects can also be due to the fact that L2 speakers sometimes “amalgamate” parsing strategies from their L1 and L2 (p. 96), as found by Hernandez, Bates, and Avila (1994) and Su (2001) and postulated by the Competition Model (Li & MacWhinney, 2012). However, in the resolution of another complex syntactic structure, namely, *wh*-dependencies, it seems that usually no evidence for L1 transfer is found (Jackson & Dussias, 2009; Marinis, Roberts, Felser, & Clahsen, 2005; Williams, Möbius, & Kim, 2001). Yet, the study by Juffs (2005) suggests that the existence of *wh*-movement in L1 poses an advantage in the correct processing of *wh*-movement in L2, which would mean that L1 does exert an influence. These mixed findings show that in the domain of syntax, transfer effects are apparently either weaker, less frequent, or more difficult to observe than, for example, in the lexical domain discussed earlier. This raises the question, if and under which circumstances L2 syntactic and morphosyntactic processing are affected by L1 transfer.

Several other studies investigating a range of other syntactic phenomena, namely the use of animacy and word order cues in sentence interpretation, have found evidence for transfer of L1 processing strategies. Liu, Bates, and Li (1992) investigated the transfer of sentence processing strategies in Chinese–English and English–Chinese bilinguals of different AoAs and proficiencies. In order to indicate the agent of an action, Chinese relies above all on animacy cues, while English relies mostly on word order cues (cf. cues in the Competition Model, section 1.5.2). The question was which cues L2 speakers would rely on. Sentences manipulating word order and animacy cues were presented auditorily in subjects´ respective L2s. Subjects had to name the agent of the sentence. Interpretation of critical stimulus sentences such as *The horse is kicking the carrot.* (word order and animacy cues

[6] Curiously, this study is seen as support for native-like L2 processing without L1 influences by some authors (e.g., by Kroll and Dussias (2006, p. 188): “The results failed to show any qualitative differences between the native and second language speakers.”, also by Dussias (2003) and Jackson and Dussias (2009), but not Barto-Sisamout et al., (2009, p. 4). However, Frenck-Mestre and Pynte (1997) explicity state: “Our results showed a localized effect of the native language, whereby readers hesitated momentarily at the region of the verb when reading in the second language if the verb behaved differently in the native language.” (p. 143) and conclude: “Lastly, we have demonstrated that the bilingual’s native language can produce a localized effect on sentence processing in the second language.” (p. 144).

coincide) vs. *The carrot is kicking the horse.* (word order and animacy cues do not coincide) was compared to the interpretation of, for example, *The horse is kicking the cow*. and **The horse the cow is kicking.*. L1 transfer effects were found in both late bilingual groups. Late English–Chinese bilinguals continued to rely heavily on word order, even if that meant identifying an inanimate object such as a carrot as the agent. Late Chinese–English bilinguals, on the other hand, relied mostly on animacy cues (and thus semantic plausibility) when interpreting English sentences. These results were replicated by Su (2001) in an offline writing task using the same language pairs. Both studies also found that transfer was mediated by proficiency, that is, transfer decreased with increasing proficiency. In an earlier study, McDonald (1987) also found that Dutch–English and English–Dutch bilingual speakers first relied on the cues preferred in their L1 (i.e., word order in English and case inflection in Dutch), but started to rely more on the preferred L2 cues with increasing proficiency. In a similar vein, Kilborn (1989) found that native German speakers and native English speakers transfer their L1 sentence interpretation preferences to L2 processing, that is, morphological cues for L1 German speakers and word order cues for L1 English speakers. All these studies are in line with the Competition Model, which predicts great reliance on L1 cues in beginners´ L2 processing and a shift towards L2 cue use at higher proficiencies. For the second experiment of the present thesis (cf. chapter 4.2), which investigates L1 gender transfer in a sentence context, this could mean that L1 gender cues are active in L2 processing, even if these cues have no importance in L2.

3.2.2 L1 transfer of morphology/morphosyntax

L1 influences in morphology/morphosyntax have been less extensively investigated than in L2 syntactic processing. In the case of morphology, transfer is a bit different. As MacWhinney (2005) states "The fact that morphosyntax is not subject to transfer is a reflection of the general Competition Model dictum that "everything that can transfer will. In the areas of phonology, lexicon, orthography, syntax, and pragmatics, there are attempts to transfer. However, in morphology there is no transfer because there is no basis for transfer. The exception here is between structurally mapable features, as in the example of gender transfer from Spanish to German." (p. 59). So instead of observing transfer of processing strategies as in the domain of syntax or transfer of L1 word meanings, research on L1 influences in the domain of morphology strongly suggests that the presence or absence of certain morphosyntactic features in L1 plays a major role in L2 acquisition and processing. Various studies have found that if a certain morphological marking is absent in L1, L2 speakers rarely show native-like processing of this feature in L2. If, however, the feature is also present in L1, L2 speakers can show native-like processing. This finding was, for example, obtained by Jiang (2004) in a self-paced reading task with Chinese learners of L2 English and a native control group. Jiang (2004) found that the Chinese learners were insensitive to number disagreement. This might have to do with the fact that Chinese, contrary to English, lacks morphological number marking. The effect was replicated by Jiang (2007). Similar results on the importance of L1 and L2 similarity for the acquisition of L2 morphology are provided by De Diego Balaguer, Sebastián-Gallés, Díaz, & Rodríguez-Fornells (2005), McDonald (2000), Montrul (2001), Slabakova (2000). The lack of a P600 (and LAN) to subject–verb agreement violations in L2 English processing of Japanese speakers (Ojima et al., 2005) and Chinese speakers (Chen et al., 2007) can probably also be explained by the fact that Japanese and Chinese lack this type of morphological agreement marking. A study by Weber and Lavric (2008), on the other hand, testing German L2 speakers of English on aspects of English verb-morphology also present in German found them indistinguishable from native speakers. As we shall see in the next section, the similarity and presence or absence of morphosyntactic aspects in L1 has also been suggested to play a role in the acquisition of grammatical gender.

As we have seen, in the domain of syntax and morphosyntax L1 influences in L2 processing have also been observed. It seems that results were somewhat inconclusive only regarding *wh*-movement and relative clause attachment. Kroll and Dussias (2006) attribute these differential results to the different types of tasks and techniques used (p. 190), whereas Frenck-Mestre (2002) emphasizes the differential influences of proficiency, language exposure, and language dominance. Another possibility for the sometimes mixed results is the role of markedness. As Ortega (2009, p. 37) points out, markedness plays an important role in transfer, since forms that are more marked in the L2 than in the L1 will pose more difficulties than vice versa. Furthermore, as mentioned above, sometimes a lack of finding L1 transfer simply points to an amalgamation of L1 and L2 strategies which leads to the development of a special learner "interlanguage" (Hernandez et al., 1994; Lee, 2001; Sasaki, 1991; Su, 2001).

Another comment regarding AoA, proficiency, and language interference is appropriate. Various studies have reported proficiency effects, which means, that usually higher proficient bilinguals experience less interference than lower proficient bilinguals. Concerning AoA, it has been observed that early bilinguals dispose of special mechanisms helping them to keep their two languages apart and to prevent interference (e.g. Bialystok, 2005, 2006). As stated by Hernandez et al. (2005) "Early, balanced simultaneous bilinguals should have the clearest evidence of language separation." (p. 222). Nevertheless, sometimes L1 interference has also been observed in early bilinguals (McDonald, 2000).

So in the section on transfer in the bilingual lexicon we saw that L1 transfer is possible, especially for nouns. In the present section, we saw that L1 transfer is also possible in the domain of syntax and morphosyntax. Since grammatical gender is a part of morphosyntax, one can ask the question whether L1 transfer is also possible for grammatical gender.

3.3 L1 grammatical gender transfer

Since grammatical gender is an aspect of grammar that is especially difficult to acquire (as discussed in chapter 2), and L1 transfer effects have been postulated as an alternative explanation to CP and AoA effects in L2 learning, L1 transfer has also been investigated in the acquisition of grammatical gender. Studies on gender transfer can be roughly divided into those looking at a) general influences from L1 (in particular, the presence or absence of a gender system) and b) those looking more specifically at gender congruency effects between L1 and L2. Besides that, a wide variety of methods has been used. Production as well as comprehension studies have been conducted, imaging techniques as well as behavioral techniques and offline as well as online tasks have been used. The application or processing of gender agreement has been studied within the NP or in more distant gender agreement relationships. Various language pairs from the same or different language families with varying similarities of gender systems have been studied. Subjects with different AoAs and proficiencies participated and as we shall see, there is also evidence for proficiency effects.

In the following, an overview of the studies belonging to either category a) or b) will be given. Studies from category a) looking at general influences from L1 on the acquisition of an L2 gender system usually focus on whether L2 speakers of an ungendered L1 experience any more difficulties in learning or using gender in a gendered L2 than L1 speakers of a gendered language. This approach is quite similar to the previously discussed studies on the influence of the presence or absence of certain morphosyntactic features in L1 on L2 learning these features (cf. section 3.2). Usually English is used as the ungendered L1. Some studies directly compare native English speakers´ performance on L2 gender agreement to the performance of native speakers of gendered L1s, others only report general difficulties of native English speakers with gender, while again others compare gender processing to number processing. Some studies also investigate the general effect of the similarity of gender systems.

3.3.1 General L1 influences

First of all, it is worth mentioning a study that provides very general evidence for the fact that L1 gender is activated during and exerts an influence in L2 processing. Scheutz and Eberhard (2004) showed that nouns that end in *-er* in L2 English, automatically activate masculine gender features in L1 German speakers of English because they are associated with the German agentive suffix *-er.* As explained by the authors "in both German and English er is a productive morpheme that can combine with a verb stem to form a noun that denotes an agent of the verb or an instrument" (p. 564). However, such nouns like *Lehrer* (*teacher*) or *Fahrer* (*drive*r) do not have grammatical gender in English but are of masculine gender in German. Apparently, L1 morphosyntactic features like the *-er* suffix can be automatically activated through the processing of a morphologically similar L2 representation. This suggests, as stated before in section 3.1 reviewing lexical transfer effects, that the activation of nouns from the L1 lexicon can have consequences on the morphosyntactic level and thus for grammatical gender processing which goes beyond mere lexical interference.

Furthermore, in various studies it has been shown that L2 learners, who are native speakers of languages which lack gender, have more difficulties in acquiring the L2 gender feature than native speakers of a language that has gender. For example, in an auditory naming task with gender-congruent and incongruent determiners used as primes, Guillelmon and Grosjean (2001) found that late English–French bilinguals were insensitive to gender markings when processing L2 French. Subjects listened to phrases of the structure determiner–adjective–noun and had to repeat the noun as quickly as possible. Only the naming times of early English–French bilinguals were influenced by the gender congruency of the determiner, compared to a baseline of a gender-neutral possessive pronoun. Apparently, the late English bilinguals used the same strategy when processing French and English, where gender-markings are absent. More evidence demonstrating the difficulties of native English speakers with L2 gender is presented in the earlier mentioned study by Franceschina (2001) (cf. section 2.1). In her case study, she observed that after 24 years of living in Spanish-speaking countries a native English speaker continued to have problems with Spanish gender agreement but made few errors in number agreement.

Also the previously mentioned study by Sabourin, Stowe, and De Haan (2006) found evidence for English speakers having greater difficulties with a gendered L2 than L2 learners of gendered L1s. Sabourin et al. (2006) conducted an offline study with pen-and-paper questionnaires in order to investigate the influence of different L1 gender systems on the learning of an L2 gender system. The L2 was Dutch, a language with an intransparent gender system. The L1s either had a gender system similar to the L2 (German), a different gender system (Romance languages) or no gender system at all (English). The authors found a main effect of L1, with German speakers performing the best and English speakers the worst. According to the authors, German subjects with an L1 similar to Dutch could rely on surface transfer, while English speakers had to learn gender markings from scratch. An important point of criticism is the fact that gender congruency between the L2 Dutch and the gendered L1s was not controlled for. As we shall see in the next section, gender congruency between languages plays an important role. So the reason why German subjects performed better than other subjects might have to do with the fact that by coincidence many items shared the same gender across German and Dutch. However, in an ERP experiment, Sabourin and Stowe (2008) found further evidence for the importance of language rule similarity in the acquisition of an L2 gender system. In their experiment testing verbal domain dependency and grammatical gender in L2 Dutch, only native German speakers and the control group showed a P600 in response to both violations. Romance L1 speakers, on the other hand, only showed a P600 in response to violations of verbal domain dependencies. The authors attribute this to the lack of similarity between the Dutch and the Romance gender system.

Sagarra and Herschensohn (2010) also provided evidence that English speakers have difficulties with L2 gender. Subjects were beginning and intermediate English L2 learners of Spanish and a native control group. Processing of noun-attributive adjective agreement violations in gender and number was examined using a GJT (offline) and a self-paced reading task (online) with comprehension questions. In the offline task, both L2 proficiency groups were highly accurate but made significantly more errors than the native group. Also contrary to the native group more errors to gender violations than to number violations were made. In the online task, the intermediate group behaved just as the native group. The beginners´ group, however, showed no sensitivity to gender and number violations contrary to the offline task. Proficiency effects were also found, as low-proficient speakers differed significantly from native speakers in the online task, while intermediate speakers did not. So this study demonstrates that for native English speakers gender agreement, which is absent in English, is more difficult than number agreement, which is present in English. This confirms other studies showing that gender agreement is more difficult than number agreement for native English speakers (Franceschina, 2001; Gillon Dowens et al., 2010; White et al., 2004). Sagarra and Herschensohn (2011a, 2011b) report similar proficiency effects in gender agreement processing of native English speakers in their L2 Spanish.

The earlier cited ERP studies (cf. section 1.3.2) by Gillon Dowens, Vergara, Barber, and Carreiras (2010) and Gillon Dowens et al. (2011) also found evidence for the influence of the presence/absence of an L1 gender system. Gender and number agreement violations in determiner–noun and noun–adjective constructions were investigated in L2 Spanish. As mentioned before, subjects were L1 speakers of English and native Spanish speakers as a control group (Gillon Dowens et al., 2010) and L1 Chinese speakers (Gillon Dowens et al., 2011). Native Spanish speakers exhibited an ELAN and a P600 in response to both types of violations. Non-native results, however, showed that the L1 English group (2010) exhibited greater difficulties with gender than number processing, as evidenced by differences in both components´ amplitude, distribution and latency, as well as higher error rates for gender agreement in sentence acceptability judgments. This is probably due to the fact that English lacks gender but not number agreement. Chinese, on the other hand, lacks both morphosyntactic features and the L1 Chinese group (Gillon Dowens et al., 2011) exhibited no ELAN to any of the constructions. This combination of results of the two studies favors an L1 transfer account over other possible explanations that were originally presented in the previous study (Gillon Dowens et al., 2010).

An especially interesting study in this regard is the study by Foucart and Frenck-Mestre (2011), who found that L1 German speakers of L2 French were only sensitive to L2 gender violations when the structure in question was also marked for gender in German. German is a language with grammatical gender but lacks gender-marking for plural forms. Gender agreement violations occurred between the determiner and the noun, between a post-posed adjective and the noun and a pre-posed adjective and the noun. For the adjective violations, the gender-unmarked plural form was used so that the gender of the nouns could not be inferred from the French determiner. The native control group exhibited a P600 in response to all agreement violations, regardless of the structure used. The L2 group, however, only revealed a P600 when the violation occurred between the determiner and the noun, but not for violations between the determiner and a pre-posed or post-posed adjective, that is, the forms presented in plural. It is very likely that this is due to L1 influences, as in German the plural form does not require gender agreement, but the lack of an effect might also be due to agreement distance (cf. section 2.2). However, in an earlier study by Frenck-Mestre, Foucart, Carrasco, & Herschensohn (2009, experiment 1), L1 German speakers did not show a P600 in response to gender agreement violations in plural forms, either. So interestingly, difficulties in processing L2 gender are not only found in native speakers of ungendered languages but also when native speakers of a gendered language encounter structures that lack gender marking in their L1.

However, it has to be noted that Frenck-Mestre et al. (2009, experiment 1) also found that native English speakers of intermediate proficiency showed a P600 in response to gender violations in French, even though it was more attenuated than the P600 exhibited by the native control group. As mentioned above, the German subjects showed no sensitivity to gender violations, which the authors attributed to the fact that the stimuli were in the plural and German does not require gender marking for plural forms. As noted by the authors, a possible explanation is that according to the Competition Model (cf. section 1.5.2), structures that are different between languages are more problematic to acquire than structures that are unique in L2, because only in the former case competition can arise (p. 96). Therefore, native-like processing of gender agreement in plural forms should be easier to acquire for English subjects, whose L1 completely lacks gender agreement, than for German subjects, whose L1 gender agreement is neutralized for plural forms.

Furthermore, other studies have shown that if the L1 lacks gender, this does not always mean that those L2 learners are necessarily completely insensitive to L2 gender agreement. White et al. (2004), for example, found effects of proficiency but no L1 influences (ungendered English vs. gendered French) in Spanish L2 gender processing. Lower proficient subjects had more difficulties with gender than number processing, but intermediate and advanced speakers did not differ in their performance from native speakers on either structure. White et al. (2004) concluded that L2 gender is acquirable by native speakers from gendered as well as ungendered languages. Nonetheless, this native-like performance of L2 speakers in gender processing could also be due to the low agreement distance and simplicity of the structure tested, that is, determiner–noun agreement of high-frequent transparent nouns.

In the also previously discussed study (cf. section 1.5.2) by Tokowicz and MacWhinney (2005), the authors found that L1 English speakers, who lack grammatical gender in their native language, were very well sensitive to determiner–noun agreement violations in their L2 Spanish despite their low proficiency. However, this sensitivity only became apparent in their online performance as evidenced by a native-like P600. This indicates native-like sensitivity to morphosyntactic violations, while behavioral performance to this type of violation was at chance. As mentioned before, a P600 was also exhibited for another structure that is similar in L1 and L2 (tense-marking violations). Yet, a P600 was not shown for violations of a structure that is different in L2 (determiner–number agreement), meaning that in this case, morphosyntactic processing was not similar to native speakers. These results led the authors to conclude that in line with the predictions made by the Competition Model, different structures are more difficult to acquire than unique structures. Therefore, it is likely that in languages where grammatical gender differs in congruency, L2 learners should experience a lot of problems. This prediction is especially important for Experiment 1 of the present thesis, which investigates gender congruency effects across German and Spanish. Studies investigating congruency effects in grammatical gender will be discussed in the next section.

Foote (2011) also reported no difference in sensitivity to gender agreement violations between early and late English-Spanish bilinguals in a moving window word-by-word sentence reading task in L2 Spanish. Sensitivity to subject–verb number agreement and noun–adjective gender agreement, as indicated by reading times, was similar across both groups when reading for comprehension. However, Foote (2011) notes that this unexpected sensitivity to gender violations of the late bilinguals could be due to the exceptionally high proficiency of their late bilinguals as almost all of them were Spanish teachers.

In summary, L1 effects in the domain of grammatical gender have been found by many but not all studies. Mainly English has been used as an ungendered language but there are also findings for Chinese. As mentioned in section 2.1 and confirmed here once again, it has been repeatedly shown that L1 speakers of English have more problems with gender agreement than with number agreement (Franceschina, 2001; Gillon Dowens et al., 2010; Sagarra & Herschensohn, 2010, (only GJT); White et al., 2004). It seems that in some cases, the successful processing of L2 gender by native English speakers can be attributed to ceiling effects, as in the case of White et al. (2004), or to the very high proficiency of the subjects as in the case of Foote (2011). In addition, the study by Tokowicz and MacWhinney (2005) demonstrated that English speakers, despite showing a P600, had a lot of difficulties with accuracy. So overall, it looks like L1 exerts an influence in L2 gender processing. This seems especially true considering the results by Scheutz and Eberhard (2004) and the finding of Foucart and Frenck-Mestre (2011) and Frenck-Mestre et al., (2009, experiment 1) that L1 speakers of German, a language with a complex gender system, are insensitive to L2 gender agreement for structures that lack gender-marking in German. Nevertheless, Frenck-Mestre et al. provided an example of native English speakers showing sensitivity to L2 gender agreement.

It is also important to note that, so far, studies have mostly looked at effects of L1 characteristics, while L2 characteristics might also exert an influence. This has been demonstrated by, for example, Frenck-Mestre et al., (2009, experiment 2), who showed that overt phonetic cues in L2 French gender agreement lead to more pronounced P600s in native speakers and L1 Spanish speakers. In the present thesis, Experiment 2 will further investigate the question whether L2 characteristics can influence L1 transfer, specifically in the case of grammatical gender. The common question whether L2 gender processing is constrained by the presence or absence of an L1 gender system is reversed in the sense that I investigated whether L1 gender transfer is constrained by the presence or absence of an L2 gender system. Furthermore, in the first experiment, the influence of transparency of the L2 gender system is also addressed.

3.3.2 Gender congruency effects

As mentioned at the beginning of the chapter, studies have been investigating grammatical gender transfer in a more general sense by looking at whether the success of L2 gender acquisition is influenced by the L1 having gender or no gender. Stated in terms of the Competition Model (Competition Model, cf. section 1.5.2), this corresponds to the acquisition of features that are unique to the L2. Other studies, however, have looked more specifically at effects of gender congruency, that is, comparing cases where L1 and L2 assign either different or the same gender values to a certain noun. The noun *apple,* for example, is feminine in Spanish (*la*$_{\text{-fem}}$ *manzana*$_{\text{-fem}}$), but masculine in German (*der*$_{\text{-masc}}$ *Apfel*$_{\text{-masc}}$). *Apple* is thus “gender-incongruent” across Spanish and German. The noun *door*, on the other hand, *la*$_{\text{-fem}}$ *puerta*$_{\text{-fem}}$ in Spanish and *die*$_{\text{-fem}}$ *Tür*$_{\text{-fem}}$ in German, is feminine in both languages and therefore “gender-congruent” across the two languages. The question asked in studies investigating gender congruency effects is whether gender congruency has an influence on the acquisition and processing of L2 words. The Competition Model would predict that features that are similar in the L1 and L2, that is, nouns that are gender-congruent, would be easier to learn and process than features that are different in L1 and L2, that is, nouns that are gender-incongruent. This is due to the competition that would be experienced in the case of gender-incongruent nouns. Furthermore, various studies previously discussed have shown that interference effects prevail in many areas of L2 processing, such as the lexicon (section 3.1), syntax (section 3.2.1), and morphosyntax (section 3.2.2). The occurrence of language interference in the lexicon especially raises the question whether there will also be interference at the level of a noun´s gender representation.

According to Costa, Kovacic, Franck, and Caramazza (2003), two hypotheses can be put forward: the “gender-integrated” (p. 182) and the “language-autonomy” (p. 183) view. According to the gender-integrated view, there is only one gender system for both languages and gender representations are shared across languages. This would mean that gender-congruent nouns of different languages share the same gender feature, but gender-incongruent nouns do not. According to the language-autonomous view, the gender systems of different languages are independent of each other and do not interact. Therefore, neither gender-congruent nor gender-incongruent nouns would share their gender features across languages. For the interpretation of experimental results, this would indicate that when gender interference is found for gender-incongruent nouns, this provides strong support for the interaction of gender information across languages and thus the “gender-integrated” view[7]. If, however, results for gender-incongruent and gender-congruent nouns do not differ, this supports the “gender-autonomous” view.

Gender interference can be defined as interference caused by the activation of non-target gender information occurring at the moment of gender retrieval of a given noun. One way of investigating the interaction vs. independence of gender systems in language production of bilinguals is through picture naming tasks, and measuring RTs and error rates. Carefully selected pictures matched across congruency conditions have to be named by the participants with NPs (either determiner + noun or adjective + noun) and/or bare nouns[8]. In studies investigating only NP or bare noun production without a baseline condition, a monolingual control group is usually included for comparison. In cases where NP and bare noun naming is measured, the bare noun condition serves as a baseline against which the congruency effects in the NP conditions are compared. Therefore, it can be assured that any differences found between gender congruency conditions are not due to differences in the experimental material but to the time needed to access a gender-marked element. In any case, if

[7] It is important to note that, as Costa et al. (2003) explain (pp. 184-185), gender interference can only occur under an activation-based account of lexical selection (Levelt et al., 1999; Levelt, 1989) but not under an account which postulates automatic retrieval of grammatical gender (Caramazza, 1997; Schiller & Caramazza, 2003). Under an activation-based account, the gender node of gender-congruent nouns accumulates more activation than gender incongruent nouns as it receives activation from both the L1 and L2 noun leading to faster gender retrieval. In the case of incongruent nouns, however, the activated gender information is conflicting therefore impairing gender retrieval.

[8] The different implications of NP vs. bare noun naming will be explained at a later point.

gender interference occurs, naming of nouns should be modified by gender congruency. Likewise, if there is no gender interference, they should be the same (relative to the baseline conditions or control group).

One of the first studies investigating the interaction of the two gender systems in bilingual language production was carried out by Costa et al. (2003). The study was conducted with high-proficient early bilinguals of many different languages, namely: Croatian–Italian, Spanish–Catalan, Catalan–Spanish and Italian–French. Subjects had to name pictures by means of L2 NPs (gender-marked determiner + noun, e.g., *la mela = the apple*). Picture names either had the same grammatical gender in L1 and L2 or a different gender. In each experiment, bilingual subjects´ performance was compared to that of a monolingual control group. In the first experiment conducted with Croatian–Italian bilinguals, no gender congruency effect was observed, neither in RTs nor in error rates. In order to rule out possible artifacts, two additional experiments using the same language pair were carried out. The picture-naming task of Croatian–Italian bilinguals was also replicated as a speeded naming task in order to reveal if possible effects were covered by long naming latencies, but still no congruency effects were found. In another modification of the task, participants were required to name pictures in L1 and L2, so as to put them in a bilingual language mode[9] in order to abet language transfer (cf. Grosjean, 1998a, 1998b). Moreover, this time, participants were required to name pictures with NPs including adjectives. This was done because the L1, in this case Croatian, lacks determiners, which might therefore prevent gender transfer in the case of determiners. Croatian adjectives, on the other hand, must agree in gender with the corresponding noun. Despite these modifications, no evidence for gender congruency effects was found. Still, a robust frequency effect was obtained in all three variations of the task, in the monolingual as well as the bilingual group. This led the authors to conclude that their design would have been robust enough to reveal possible congruency effects.

Croatian and Italian are members of different language families (Slavic and Romance, respectively) and their gender systems are asymmetric, with Croatian having a three-way gender system and Italian a two-way gender system. Because of that, the failure to find a gender congruency effect might also be attributable to a lack of similarity between the languages or gender systems. Therefore, the picture-naming experiment was replicated in two more experiments with language pairs of the same language family (Romance) that have completely symmetric systems, namely, Spanish–Catalan, Catalan–Spanish and Italian–French. All these Romance languages have a two-way gender system (masculine, feminine). However, also the replication with these language pairs and an additional replication with Italian–French bilinguals failed to reveal a gender congruency effect, in RTs as well as error rates. In summary, no gender congruency effect was observed in any of the experiments and with none of the language pairs. The authors conclude that “[…] the gender values of the words in the non-response language do not affect performance in the response language. Therefore, at this point it seems reasonable to conclude that the gender properties of one language do not affect gender processing in the other language.” (p. 194).

However, the failure to find a gender congruency effect can also be attributed to other reasons. First of all, as also noted by the authors, participants were early and therefore very high-proficient bilinguals. As mentioned in previous sections of the present chapter, AoA and proficiency are known to affect language transfer. The authors themselves acknowledge that “It is possible that the degree of language autonomy of the two gender systems of a bilingual speaker depends on the degree of L2

[9] Language mode denominates “[…] the state of activation of the bilingual’s languages and language processing mechanisms at a given point in time.” (Grosjean, 1999, p. 3).

[sic]. It may be the case that the less proficient a bilingual speaker is, the greater the interaction between the gender systems. Future research is needed to address the impact of these variables, among others, concerning the role of the non-response language during speech production." (p. 194). Second, the number of subjects in three of the five bilingual groups tested was very low and might have been too small to reveal gender congruency effects.

Third, as also pointed out by Salamoura and Williams (2007b, p. 259), another characteristic of the language pairs used in the present study is that (gender-transparent) Romance languages usually also fail to show the "classic" monolingual gender interference effect in NP production, which could be the reason for the failure of finding a bilingual gender interference effect. The classic monolingual gender interference effect has been investigated in different languages, mainly using a picture-word-interference (PWI) paradigm. In the PWI paradigm, participants are shown a picture, which they have to name by means of a gender-marked NP while ignoring a distractor noun printed on the picture. For some languages, such as Dutch (Schriefers, 1993) and German (Schriefers & Teruel, 2000), gender congruency between the picture name and distractor word has been shown to affect RTs: Longer RTs are observed in the case of gender-incongruent picture–distractor pairs and shorter RTs in the case of congruent pairs. However, this is not the case with Romance languages (Alario & Caramazza, 2002 (French); Costa, Sebastián-Gallés, Miozzo, & Caramazza, 1999 (Catalan and Spanish); Cubelli, Lotto, Paolieri, Girelli, & Job, 2005 (Italian); Miozzo & Caramazza, 1999 (Italian); Miozzo, Costa, & Caramazza, 2002 (Italian and Spanish)).

An explanation for these differential findings between Romance and Germanic languages concerning NP production with determiners has been put forward by Miozzo and Caramazza (1999), termed the "early selection" and "late selection hypothesis". Germanic languages are "early selection" languages since the determiner form can be specified relatively early in the speech production process because the determiner form is mostly determined by the grammatical gender of the noun. In Romance languages, however, the correct determiner form can only be selected relatively late in the speech production process because the phonology of the noun onset (vowel vs. consonant) also plays a role for selecting the correct determiner form. Therefore, it is hypothesized that by the time the determiner is selected, a potentially occurring competition at the level of gender selection has probably already ceased. Importantly, this means that the possibility of finding a gender interference effect in NP production is not precluded for Romance languages. It might just be more difficult to reveal. Nevertheless, since so far it has not been possible to detect a gender interference effect for Romance languages in monolingual NP production, it is probably not surprising that the effect was also not found in a bilingual context by Costa et al. (2003).

Nonetheless, it is important to note that more recently, a monolingual gender interference effect has been found for Romance languages with the PWI paradigm in bare noun production (Cubelli et al., 2005 (Italian); Paolieri, Lotto, et al., 2010 (Italian and Spanish); Paolieri et al., 2011 (Italian))[10]. This variation of the classic monolingual gender interference effect differs from the classic effect in three aspects: First, the effect is only observed in bare noun naming and disappears in NP production (Cubelli et al., 2005) which is in line with the previously discussed results. Second, the effect is reversed in the sense that for congruent picture–distractor pairs, longer instead of shorter naming times than for incongruent pairs are obtained. Third, the monolingual gender interference effect in

[10] A monolingual gender interference effect has also been found with other slightly different methods such as a PWI with a determiner as the distractor word (Alario, Ayora, Costa, & Melinger, 2008) or determiner primed picture naming (Alario, Matos, & Segui, 2004).

bare noun production has not been found in Germanic languages so far (Dutch: La Heij, Mak, Sander, & Willeboordse, 1998, experiment 2). These differences suggest that other mechanisms than in the classic effect are at work. As Cubelli et al. note "[…] the most prominent current models of language production cannot easily account for the effect of grammatical gender in the production of bare nouns." (p. 49). So in order to explain these differences from the classic interference effect, Cubelli et al. and Paolieri et al. (2011) had to make several new assumptions that are not always compatible with current language production models.

First, Cubelli et al. (2005), Paolieri et al. (2011) and Paolieri, Lotto, et al. (2010) observed an effect in Italian and Spanish bare noun production. As stated by Paolieri, Cubelli, et al. (2010, p. 2), the gender of a noun is always selected upon noun retrieval ("lexical hypothesis"), and not only in NP production ("syntactic hypothesis")[11]. According to the lexical hypothesis, grammatical gender is retrieved upon lexical access so that gender congruency effects are also visible in bare noun production. According to the syntactic hypothesis, which is entertained by the most important language production models, gender congruency effects are only observed in NP production, because grammatical gender information solely becomes available when needed to compute agreement (Caramazza & Miozzo, 1997; Levelt et al., 1999). Paolieri, Cubelli, et al. (2010) and Cubelli et al. (2005) favor the lexical hypothesis over the syntactic hypothesis, at least for languages like Italian (see next argument). Second, Cubelli et al. point out that in Romance languages with gender transparent noun endings, such as Italian and Spanish, the morphological structure (noun ending) and thus phonological form of nouns is determined by the noun´s grammatical gender (p. 46). Therefore, the phonological form of a (bare) noun cannot be determined without accessing its gender information. Cubelli et al. (p. 53) propose a "double selection mechanism" where prior to determining a noun´s morpho-phonological form, semantic and syntactic information are selected in two steps, first the noun stem via the semantic information, then the nominal ending via the gender information. Thus, a bare noun effect is only found in languages that are morphologically marked for gender. This would explain why a bare noun effect is found in Romance languages with transparent gender markings, but not in Germanic languages. Third, naming times for congruent nouns are slower than for incongruent nouns because gender-congruent nouns accessing the same gender information receive more activation and compete for selection[12]. This mechanism is similar to the explanation proposed for semantic interference effects in which longer RTs for semantically related items are also obtained (Cubelli et al., pp. 43 and 53). According to the double selection model, the monolingual gender interference effect in bare noun production is due to interference at the time of selection of noun inflection.

In addition, regarding the gender interference effect in NP production, some authors assume that interference observed at the level of determiner selection reflects gender interference because determiner selection requires gender retrieval (Levelt et al., 1999; Schriefers & Jescheniak, 1999), other authors have argued that gender interference in NP production reflects competition in determiner retrieval rather than at the level of gender feature selection (Alario & Caramazza, 2002; Alario, Matos, & Segui, 2004; Caramazza & Miozzo, 1997). However, note that in a bilingual context, the explanation with interference at the level of determiner selection is no longer feasible. Here, the

[11] It is important to point out that in a bilingual context, both hypotheses can be entertained within the gender-integrated hypothesis (Paolieri, Lotto, et al., 2010., p. 3).

[12] Note that such a mechanism would need a bidirectional link between noun and gender node. According to the most prominent speech production models, however, the link is unidirectional, with no activation sent from the gender node (Schriefers & Jescheniak, 1999, p. 579).

gender feature has to be activated in order for gender interference between the two languages to occur.

Evidently, there is still a discussion going on regarding the exact locus of the gender interference effects in different paradigms and phrase types. As the present thesis deals with gender interference in the bilingual language system, a further discussion of the locus of gender interference in the different language production models is beyond scope. For the present discussion, it is most relevant to point out that, as stated above, the reason why Costa et al. (2003) did not find a bilingual gender interference effect might be because in a monolingual context, gender interference is not found in NPs for Romance languages. When investigating gender interference in the bilingual context, it could be important to take into account whether and under which circumstances the chosen languages show gender interference in a monolingual context. The present results suggest that the decision whether to investigate gender interference by means of NP production or bare noun production has to be considered carefully. For now, we can conclude that, ideally, for Germanic languages, interference effects should be investigated with NPs, while for transparent Romance languages, it would probably be better to investigate gender interference with bare nouns.

So even though Costa et al. (2003) found no evidence for gender transfer and preclude the possibility of the interaction of the two gender systems of a bilingual, several characteristics of the study give room for alternative explanations. As we shall see in the following, other studies specifically set out to manipulate the variables which were possibly responsible for the lack of finding an effect. These studies used language pairs from other language families (that also show the monolingual gender effect). Their participants were late and lower proficient bilinguals, and they included more participants. In most cases, these studies did find evidence for gender transfer. Besides picture naming, translation tasks and lexical decision tasks were also used. In the following, an overview of the studies investigating gender congruency effects will be given. First, I will review studies investigating gender interference effects in picture naming in NPs between non-Romance languages. Then I will review studies investigating interference effects between non-Romance languages in translation studies. Last, I will review some recent studies investigating gender interference effects in picture naming and a translation task with Romance languages.

Picture naming studies using NPs and non-Romance languages

Bordag and Pechmann (2007) conducted four picture naming experiments with L1 Czech speakers of L2 German. No control group was tested. They found effects of gender congruency and thus evidence for a shared bilingual gender system, contradicting Costa et al. (2003). Czech is a Slavic language and German a Germanic language, but both have a three-way (masculine, feminine, neuter) and therefore symmetric gender system. Pictures had to be named by means of noun + adjective NPs and bare nouns. Participants had intermediate and upper-intermediate proficiency, as the authors assumed that L1 gender interference would most likely occur at this proficiency level (p. 301). Several other aspects known to influence L2 (gender) processing, such as language mode and noun ending transparency, were also investigated. Experiments 1 and 2 manipulated subjects´ language mode. In experiment 1, subjects had to name pictures in L1 as well as L2 and were thus close to a bilingual language mode. In experiment 2, subjects were brought closer to a monolingual mode, as L2 was the only response language. In experiment 3, the effect of noun ending transparency was investigated in addition to gender congruency. Three groups of noun endings were investigated: Group A contained L2 nouns with a gender-typical and thus transparent termination, Group B contained L2 nouns with a gender-ambiguous termination, and Group C contained L2 nouns with a gender-atypical termination.

In experiment 4, the effect of noun transparency was investigated in comprehension, a GJT, without investigating gender congruency. The following results were obtained: In experiment 1 and 2, a gender congruency effect, that is, faster naming times for congruent pictures, was found in the RTs under both the bare noun[13] and the adjective condition but not in the error rates. Therefore, the authors concluded that L2 speakers are not able to reduce L1 activation even when they know the response language beforehand (experiment 1) and neither when they are close to a monolingual mode (experiment 2). This is in agreement with the Competition Model assumption that all languages known by a speaker are always activated and compete for selection. In experiment 3, a gender congruency effect (only RTs) in bare nouns and NPs as well as an effect of noun ending transparency (RTs and error rates) was found. In the GJT of experiment 4, an effect of L2 noun endings[14] was also found, in RTs as well as error rates. Thus, the gender congruency effect was robustly shown in three production experiments, therefore challenging the results of Costa et al. (2003).

Another study, conducted by Lemhöfer, Spalek, and Schriefers (2008), investigated the organization of the L2 gender system with different language pairs. Gender interference was investigated in production through a PNT and also in comprehension, in an LDT. (This study is especially important for the present thesis, since my first experiment is to a large extent a replication with German–Spanish and Spanish–German bilinguals.) L1 speakers of German were tested in their L2 Dutch. Participants were of intermediate proficiency. Dutch and German are both members of the Germanic language family but have asymmetric gender systems, as Dutch has a two-way gender system (common gender including feminine and masculine gender and neuter gender) but German has a three-way gender system[15]. The monolingual gender interference effect has been observed for both languages. Experiment 1 was an LDT. Nouns were primed by definite determiners, which are gender-marked in Dutch, or indefinite determiners, which are gender-neutral in Dutch. Experiment 2 was a PNT using the same material. Pictures were named as bare nouns and with NPs including the gender-marked definite determiner. Phrase Type conditions were blocked. Experiment 3 was a replication of experiment 2, but a training session was included in order to lower error rates and render the data more reliable. Cognate items as well as noncognate items were used. Cognates are words that share

[13] In this study, no control group was included. However, a gender interference effect in bare noun production was consistently found across experiments 1, 2, and 3 “with different subjects, modified designs and different items” (p. 305). It has to be noted, though, that contrary to the previously discussed monolingual bare noun interference effect, the bare noun effect found in this study points in the same direction as the NP effect, i.e. faster naming times for congruent nouns. Bordag and Pechmann explain that contrary to the monolingual PWI experiments, in bilingual gender interference experiments, the lemmas of the L1 and L2 translation equivalents share the same underlying concept which might favor the competition of their grammatical features. This could potentially make gender interference and facilitation effects also observable in bare noun naming. Meanwhile, in the case of different underlying concepts, gender information only competes, if necessary, for agreement computations (pp. 305-306).

[14] The effects of noun transparency are interpreted by the authors as providing additional support (in addition to gender interference effect) for activation-based accounts as the noun ending increases activation for transparent nouns. These effects are also interpreted as providing support for cascading models (e.g., Interactive Activation Model of Dell (1986), Independent Network Model of Caramazza (1997)) rather than serial models (Levelt et al., 1999; Levelt, 1989) because the phonological information of a noun can influence gender access. According to the authors, following this line of reasoning support for activation-based cascading models would also be provided by the later discussed study by Lemhöfer, Spalek, and Schriefers (2008) because they found an effect of Cognate Status.

[15] Note that, since German and Dutch have asymmetric gender systems, with the German masculine and feminine genders both mapping onto Dutch common gender, Lemhöfer et al. (2008) use the term “gender compatibility” rather than “gender congruency”.

the same meaning and are form-similar across languages. In many studies, cross-language influences have been shown to be stronger for cognates than for noncognates (Lemhöfer et al., 2008, p. 314). The following results were obtained: In experiment 1, gender interference effects were found, and these effects were stronger for cognates than for noncognates. This was true for RTs (only significant by participants) as well as for error rates. Overall, as expected, cognates were also recognized faster and fewer errors were made. The authors concluded that the two grammatical gender systems of a bilingual interact during visual word recognition. For experiment 2, no significant effects of gender interference were found in the RTs, neither for cognates nor for noncognates, which was possibly due to a loss of power because of high error rates. Cognates were named faster (only significant by participants) than noncognates. Overall, voice onset for bare nouns was faster than for NPs. In error rates, gender interference effects were found, i.e., more errors were committed in the incongruent than in the congruent condition. Word form similarity had no influence on error rates. Fewer errors were made in the bare noun condition. An offline gender assignment task was also conducted and the error pattern mirrored that of the error pattern of the online task: Gender congruency effects were found and these effects were stronger for cognates than for noncognates. In experiment 3, Lemhöfer et al. (2008) investigated the representational stability of gender representations by analyzing the consistency of assignments across various repetitions. According to the consistency of gender assignments, items were divided into “stable” and “unstable” items. Because of the training session, error rates were greatly reduced, and this time, a gender congruency effect was obtained not only in error rates but also in RTs (significant by participants and items for cognates and only by participants for noncognates). In addition, it was found that assignments of gender-incongruent nouns were more inconsistent than those of gender-congruent nouns. Furthermore, in RTs, the gender congruency effect was only significant for unstable items. Regarding error rates, there was a congruency effect for stable as well as unstable items. Thus, the authors concluded that the gender congruency effect found in experiment 2 might actually stem from unstable and incorrect gender representations due to transfer effects in the acquisition phrase, rather than online competition (“acquisition-based account” vs. “online account”, respectively, p. 326). However, they did not want to completely rule out the possibility for online competition. The authors also considered the fact that L2 Dutch has an intransparent gender system important. They assume that L1 influences might play a greater role when L2 has an intransparent gender system compared to a transparent one, as in the case of many Romance languages (pp. 327-328).

In a follow-up study, Lemhöfer, Schriefers, and Hanique (2010) further investigated the issue of representational stability (operationalized as response consistency across item repetitions and self-rated gender certainty), as well as the effects of training and feedback, which will not be discussed here. The same material as in the previous PNT was used and the positive effects of training on accuracy were replicated as well as the gender congruency effect, which was also stronger for cognates. Again, gender-incongruent items proved to be more unstable than gender-congruent items, as indicated by response consistency across repetitions as well as gender certainty ratings. Furthermore, in the offline gender assignment task administered in the familiarization phase, more errors were committed for gender-incongruent noun than for congruent nouns, and this effect was more pronounced for cognates than noncognates. These findings concerning response consistency and accuracy of offline gender assignment raise an important alternative explanation for the alleged online interference effects: The high error rate for incongruent L2 nouns might actually stem from incorrect or unstable representations rather than online competition effects. It is possible that if L2 learners use their L1 gender system as a point of departure in L2 acquisition, it is simply more difficult and unlikely to acquire correct and stable representations for incongruent nouns than for congruent nouns in the first place. According to Lemhöfer et al. (2008) such an “acquisition-based” account would also question the interpretation of online interference effects. Unfortunately, to my knowledge, this issue has not been addressed any more in subsequent experiments investigating gender congruency effects. In the first experiment of the present thesis, an offline gender assignment task is conducted in addition to the two online experiments (PNT and LDT) in order to address the possibility of gender congruency effects stemming from incorrect gender representations rather than online competition. By comparing the amount of errors rates in the offline task to online interference effects in RTs and error rates across conditions, it can be estimated whether potential online processing difficulties with incongruent items are due to a lack of knowledge, due to erroneous acquisition, or rather due to online interference effects.

Translation task studies using NPs and non-Romance languages

Sometimes another type of production task is used to investigate gender interference, namely, a translation task. As pointed out by Bordag and Pechmann (2008, pp. 140 and 142), translation tasks differ from production tasks in a number of aspects: First of all, in translation tasks it seems obvious that both the L1 and L2 gender node will be activated, but it is possible that they are not activated simultaneously, thus rendering gender interference unlikely. Second, in translation tasks, the L1 word to-be-translated is thought to provide a cue regarding the target language (“use L2, not L1”). Such a cue might reduce cross-language competition and is absent in picture naming tasks. As a consequence, different results in cross-lingual tasks are often obtained with these two tasks, and these differences might render important insights about language-processing in bilinguals. In the following, two translation tasks investigating gender interference in non-Romance languages, using NPs, will be discussed.

A translation task from L1 Greek to L2 German was carried out by Salamoura and Williams (2007b), who tried in many ways to optimize conditions for finding a gender congruency effect and explicitly addressed the weak points of the Costa et al. (2003) study. Greek and German are from different language families (Greek and Germanic) but have symmetric gender systems, that is, both languages have a three-way gender system with the same gender values (masculine, feminine, neuter). Furthermore, Greek and German both yield the L1 gender congruency effect (p. 261). Just as the Romance languages used in the study by Costa et al. (2003), Greek has a fairly transparent gender system (Salamoura & Williams, p. 260). As mentioned before, the German gender system, on the other hand, is much less transparent than the systems of most Romance languages. That is, contrary to Romance languages, in many cases it is not possible to derive a noun´s gender from its termination. The authors suggest that the lack of finding a gender congruency effect in Costa et al.´s study might be due to the morpho-phonological transparency of the Romance gender systems inviting more superficial gender processing (p. 268). Therefore, in a language with a fairly intransparent gender system like German, L1 gender transfer effects might be easier to obtain. Participants were required to translate adjective + noun NPs as well as bare nouns which served as a baseline. In addition, effects of word form similarity were investigated by comparing cognates and noncognates, similar to the previously discussed study by Lemhöfer, Spalek, and Schriefers (2008). A translation advantage for gender-congruent nouns relative to gender-incongruent nouns was found, but only in the case of adjective + noun NPs. When bare nouns were translated, no congruency effect was observed. This effect, as the authors interpreted, indicated the operation of an “economy principle” (p. 267), meaning that syntactic features are only selected if needed for correct production of the item in question. Which syntactic features are needed for accessing the correct morpho-phonological form of an item might depend on certain language characteristics, as explained by Cubelli et al. (2005). Cognates were translated faster but the congruency effect was similar in size for noncognates and cognates. However, more errors were made in the case of gender-incongruent cognates than for gender-congruent cognates, which was not the case for noncognates. Furthermore, no transparency effect with regard to noun terminations was found, but few transparent nouns had been included in the study. Overall, it can be said that the results of Salamoura and Williams provide support for a shared gender system between L1 and L2 and are in opposition to the results found by Costa et al.

In a later study, Bordag and Pechmann (2008) tried to replicate the gender interference effect of L1 Czech to L2 German found in their 2007 study using PNTs, this time with a translation task instead. Participants were asked to translate adjective + noun NPs and bare nouns. Phrases appeared on the screen in Czech. The adjectives to be used (*big, small*) were elicited by means of a big or small dot on the screen. In experiment 1, no gender interference in RTs was found, but it did occur in error rates. In experiment 2, no gender interference was found, neither in RTs nor error rates. However, a transparency effect was found. Since the adjectives to be used were indicated by a small or a big dot, the authors reasoned that the task actually did not constitute a real translation task, which could have prevented lexical selection of an L1 word and thus competition through L1 gender. Therefore, experiment 3 was conducted as a "true" translation task. This time adjectives were also given in the L1, instead of cueing them with small or big dots. However, again, no gender congruency effect was found, while the transparency effect of experiment 2 was replicated. The authors consider different explanations as to why the previously obtained gender interference effects by Bordag and Pechmann (2007) could not be replicated in the translation task. The explanation they regard the most plausible is based on the order of activation of L1 and L2 lemmas in picture naming vs. forward translation. They state that in picture naming, the conceptual level, L1 and L2 lemmas, and therefore also L1 and L2 gender nodes, are activated in parallel, and as a result, give rise to interference. On the other hand, with translation tasks, L1 word forms and lemmas are activated *before* L2 lemmas so that activation of L1 and L2 gender nodes is not simultaneous, making gender interference less likely. As stated above, according to Bordag and Pechmann (2008) it is also possible that the word to-be-translated provides a strong cue to the target language, lowering language competition, while concept-mediated picture naming is likely to activate both languages to an equal level. Furthermore, another explanation is provided by Paolieri, Cubelli, et al. (2010). These authors argue that the reason why Bordag and Pechmann (2008) might have failed to find an effect could be that in their task many different response types were required (pp. 3-4). Therefore, it might have taken subjects too much time to make a decision regarding the correct response so that any gender congruency effects might have disappeared by the time of response output (floor effect).

As we have seen, after the study by Costa et al. (2003) supporting the language autonomy view, various studies testing less proficient, late bilinguals and using non-Romance languages (that also yield a monolingual gender congruency effect in NPs), have found gender interference effects, thus supporting the gender-integrated view. Interestingly, more recently conducted studies also found a bilingual gender interference effect in Romance languages. These studies, including a PNT and a translation task, will be discussed in the following.

Picture naming/translation task studies using bare nouns and NPs and Romance languages

Also in Romance languages, gender congruency effects were again investigated. For instance, Paolieri, Cubelli, et al. (2010) tested Italian–Spanish bilinguals in a picture naming task (experiment 1 and 2) and a word translation task (experiment 3). In experiment 1, bare nouns and NPs (determiner + noun) had to be produced in Spanish. In experiment 2, performance in bare noun naming of the bilingual group was compared to a monolingual control group. In experiment 3, bare nouns and NPs had to be translated from L1 to L2. Clear gender congruency effects were found for the RTs (but not error rates) of the bilingual group across all three experiments and in both phrase types. Moreover, congruency effects of equal magnitude were obtained for both phrase types, which are again interpreted as evidence in favor of the "lexical hypothesis". However, contrary to the bare noun monolingual gender interference effect in the PWI paradigm, this time RTs were faster for gender-congruent than gender-incongruent items. Cubelli and Paolieri (2008) explain that these differential interference effects in the two paradigms can be attributed to the different task demands: "In L2 picture naming, the lexical representation of L1 noun spreads activation to L2 name, thus facilitating the response in the gender-congruent condition in all gendered languages; on the contrary, in the picture-word paradigm, the lexical representation of the distracter noun has to be inhibited, thus increasing response times in the gender-congruent condition in languages with inflected nouns" (p. 13). Contrary to Costa et al. (2003), this experiment showed that bilingual gender interference effects could also be obtained for Romance languages, even in NP production. The authors further claim that Lemhöfer, Spalek, and Schriefers (2008) also found gender congruency effects in bare noun production (p. 3). At first sight, this might seem to be true: There was a significant effect of Gender Compatibility in experiment 3, a PNT. In addition, the RTs (as shown in table 10 on p. 324) were 906 ms and 983 ms in the compatible and incompatible bare noun cognate conditions, respectively, and 970 ms and 1020 ms in the compatible and incompatible bare noun cognate conditions. Nevertheless, it has to be taken into account that Lemhöfer et al. (2008) did not include a native control group, so their bare noun effects could also be due to differences between the experimental conditions. The effects found for NPs, however, can be considered valid because they were compared to the bare noun conditions.

In a later study, the bilingual gender interference effect for Romance languages in bare noun naming was replicated by Morales, Paolieri, and Bajo (2011). They also found gender congruency effects in a picture naming experiment with Italian–Spanish bilinguals. No monolingual control group was included. Naming times in the incongruent condition were slower than in the congruent condition. They furthermore found evidence for suppression of L1 gender. Participants named pictures in L2 by means of a bare noun. Half the pictures were named just once and the other half, five times. After that, participants had to produce the L1 articles in response to the pictures. L1 article production times were slower for incongruent picture names than for congruent picture names. This effect was stronger for pictures that had been repeated five times than for pictures that had been named just once, thus showing an accumulation of inhibition. New pictures that had not been shown before showed no congruency effect. These congruency and inhibition effects shown in the second task provide evidence that the L1 gender information has to be actively suppressed in an L2 naming task in order to prevent interference. This, once more, points to a shared gender system for L1 and L2 rather than two autonomous systems.

Chapter summary

A few years ago, the picture regarding bilingual gender interference looked still inconclusive. However, by now, the results of the majority of the studies point to an interaction of bilingual gender systems. Even for gender transparent Romance languages, gender interference effects have been found. Thus, in the present literature overview, the study by Costa et al. (2003) is the only study that failed to find evidence for gender transfer[16]. However, it also has to be noted that an important caveat in any literature review on gender transfer studies is that it might be difficult for studies not finding a gender interference effect to get published, making it difficult to get a real overview on the number of studies that found no evidence for gender transfer.

Furthermore, considering what is known about monolingual gender interference effects, it now seems that one of the reasons why Costa et al. (2003) did not find a gender interference effect is probably due to the fact that they used NP naming in order to investigate interference effects in Romance languages (and Croatian) even though Romance languages also fail to show the classic monolingual gender interference effect in NP production. As mentioned before, another reason might be that they tested few participants, leading to a lack of statistical power, and only very high-proficient and early bilinguals. It is likely that AoA and language proficiency have a strong effect on language transfer (cf. sections 3.1 and 3.2) and therefore, possibly also on the organization of the bilingual gender system (integrated vs. autonomous). For this reason, the two experiments that will be reported in the present thesis specifically investigate proficiency effects in gender transfer (in late bilinguals). This is especially interesting since, as discussed before (cf. section 1.4), there seems to be a trade-off between AoA and proficiency, and in some cases proficiency can even compensate for the negative consequences of a high AoA in various aspects of L2 learning. To this point, proficiency effects in gender transfer have, to my knowledge, not yet been investigated.

Moreover, so far we have only seen studies that investigated gender transfer effects between languages from the same language families and with asymmetric gender systems (Lemhöfer, Schriefers, & Hanique, 2010; Lemhöfer, Spalek, & Schriefers, 2008, both L1 German–L2 Dutch) and between languages from different language families but with symmetric gender systems (Bordag & Pechmann, 2007, L1 Czech–L2 German; Salamoura & Williams, 2007b, L1 German–L2 Greek). This leaves the question open, whether transfer is also possible between languages from different language families with an asymmetric gender system. In addition, the studies discussed earlier reported only gender interference between two non-Romance or two Romance languages. So another question that remains to be answered and will be addressed in the first experiment is whether gender interference – and thus integration of gender systems – can occur between Germanic and Romance languages. As explained earlier, Germanic languages usually have a fairly intransparent gender system, while most Romance languages have a fairly transparent gender system. As mentioned by the authors of some of the studies discussed earlier (Bordag & Pechmann, 2007; Lemhöfer et al., 2008; Salamoura & Williams, 2007b), factors such as language similarity, symmetry of gender systems, and L2 noun transparency might also mediate gender interference effects. Hence, it will be interesting to see if gender interference can be found at all between Germanic and Romance languages and whether this gender interference will occur in both directions. It is also unknown whether and which of these potential interference effects will manifest themselves in the processing of NPs, in bare nouns, or in both.

[16] Except for the study by Bordag and Pechmann (2008) but, as explained above, here the failure to find an effect might be attributable to characteristics of the task (cf. Paolieri, Cubelli, et al., 2010, p. 3).

In addition, the influence of the complexity or transparency of the L2 gender system has so far, at least to my knowledge, not received much attention in the research literature. To date, the influence of L1 characteristics[17] has been mainly considered in L2 gender transfer research (such as the presence or absence of gender in L1, cf. section 3.3.1). But it is a relevant question whether transfer can also be mediated by L2 characteristics, such as transparency/opacity or simplicity/complexity of the L2 gender system. Therefore, it will be interesting to see if gender transfer will be more visible in L2 Spanish, with a transparent gender system, or in L2 German, with an intransparent gender system. A related question is whether gender transfer can also occur into an L2 that lacks grammatical gender, such as English. This question will be investigated in the second experiment reported in this thesis (cf. chapter 5). So the investigation of the influence of L2 characteristics on gender transfer is an important contribution to present research on gender transfer effects. Other factors known to influence performance on L2 tasks in general (discussed in section 2.2), are the type of processing required such as online vs. offline processing and production vs. comprehension as well as the manipulated syntactic structure and agreement distance. The influence of these factors will also be investigated in the present thesis.

[17] Except for Bordag and Pechmann (2007) and Salamoura and Williams (2007) who also investigated if gender interference effects were mediated by L2 noun ending transparency.

4. Experiment 1

4.1 Introduction

In the first experiment, an offline gender assignment task as well as an LDT and a PNT were conducted. As mentioned above, by comparing online interference effects to offline gender transfer effects it will be possible to estimate whether gender errors are committed because of transfer effects during the acquisition phase resulting in faulty representations or because of actual online gender interference. Spanish speakers were tested in their L2, German, which shows the classic monolingual gender effect in NPs. In addition, German speakers were tested in their L2, Spanish, which shows a monolingual gender effect in bare nouns. Monolingual control groups are also tested, which (as mentioned in section 3.3.2) allows for the analysis of effects in bare noun processing. Regarding the direction of gender transfer, there are various possibilities: It is possible that there will be only gender transfer from L1 Spanish to L2 German or only from L1 German to L2 Spanish, or transfer could occur in both directions. Furthermore, the effects might be visible only in NPs for one language pair and only in bare noun naming for the other language pair, or in both syntactic structures for both language pairs. Subjects were late bilinguals with different proficiency levels, ranging from low-proficient to high-proficient in order to investigate whether gender transfer effects are mediated by proficiency. Considering the literature discussed in previous sections of the present chapter, it is predicted that transfer effects are stronger at lower proficiency levels than at higher proficiency levels. For each proficiency level within each language pair, at least 20 subjects were tested in order to assure that statistical power was strong enough for the detection of gender interference effects[1]. RTs were measured in the LDT and PNT and error rates were measured in all three tasks. Further details are given in the Method section (section 4.2). Moreover, since transfer effects have been shown to be stronger for cognates than for noncognates (Lemhöfer, Schriefers, & Hanique, 2010; Lemhöfer, Spalek, & Schriefers, 2008; Salamoura & Williams, 2007b), noncognates as well as cognates were included as stimuli.

In the following, the gender systems of the two languages used in Experiment 1, German and Spanish, will be described in more detail. After the description of the German and Spanish gender systems, the predictions of Experiment 1 will be explained.

4.1.1 The German and Spanish gender systems

Like most Romance languages, Spanish has a two-way[2] and fairly transparent gender system (Costa et al., 2003; Paolieri, Cubelli, et al., 2010). Nouns can be either masculine or feminine. As a semantic rule, nouns which invariably denote one male human being are masculine, and nouns which invariably denote one female human being are feminine (Bergen, 1978, p. 869; Vera Morales, 2013,

[1] For comparison: Paolieri, Cubelli, et al. (2010) tested 12 bilingual participants in each of their three experiments. In the study by Lemhöfer et al. (2008), 20, 18, and 22 bilingual subjects entered data analysis across the three experiments, respectively. Salamoura and Williams (2007b) tested 18 bilingual participants. Bordag and Pechmann (2007) tested 18 participants in experiment 1 and 3. The number of participants in experiment 2 is unclear. Consequently, including 20 participants per proficiency group seemed a reasonably high enough number to be able to observe transfer effects.

[2] Since historically, Romance had three genders, there are some remnants of neuter gender found, e.g., in the neuter pronouns *ello*, *lo* (also an article), *esto*, *eso,* and *aquello*. Nowadays, there are no nouns with neuter gender (Corbett, 1991, pp. 214-215; de Bruyne, 2002, p. 69). The only exception might be nominalizations of adjectives and verbs which receive the neuter article *lo*: bueno (*good*)–*lo bueno*, *malo* (*bad*)–*lo malo*, *comprar* (*to buy*)–*lo comprado*, *dibujar* (*to draw*)–*lo dibujado*.

p. 3). Furthermore, even though “for non-human nouns, gender is completely arbitrary” (Franceschina, 2005, p. 95) there are morpho-phonologically transparent gender cues with a high reliability. Bergen (1978), for example, maintained that “the bases for the gender of Spanish nouns are mainly phonemic” (p. 868). In his analysis he was able to deduce a small number of rules (two semantic rules and eight phonemic rules), which allow determining the gender of 97.3 % of all nouns in the investigated data set of 38,233 nouns. The fact that it is possible to describe Spanish gender assignment with a relatively small set of rules with only a few exceptions demonstrates quite well how simple the Spanish gender system is. The most important and easiest rule, which is cited as the first rule in many Spanish grammar books (Böhringer, 2000, p. 5; Moriena & Genschow, 2010, p. 63; Reumuth & Winkelmann, 1997, p. 19), is that most nouns ending in *-o* are masculine and most nouns ending in *-a* are feminine[3]. Besides this most salient rule, other phonemes and morphemes that correlate with grammatical gender can be found. Further typically feminine endings are, for example, *-d*, -z, *-ión*, *-umbre*, *-ie*, *-is* (unstressed), and typically masculine endings are *-l*, *-n*, *-r*, *-s, -e*[4] (Bergen, 1978, p. 180). Furthermore, animals are oftentimes not specified for sex but have a default gender[5] (Franceschina, 2005). For all elements in noun phrase, gender agreement is obligatory. This means that the gender of definite (*el*$_{\text{-masc}}$/*la*$_{\text{-fem}}$) and indefinite (*un*$_{\text{-masc}}$ /*una*$_{\text{-fem}}$) articles, demonstratives, and attributive and predicative adjectives is determined by the gender of the respective noun (the same applies to grammatical number). Even though acquiring the grammatical gender of Spanish nouns seems to be one of the greatest problems for native English speakers studying Spanish as an L2 (Bergen, 1978, p. 865), acquisition of the Spanish gender system should be considerably easier than learning the German gender system which, as we shall see in the following, is much more complex[6].

German is a Germanic language and has a much more intransparent and complex three-way gender system. Nouns can be masculine, feminine, or neuter. The definite articles are *der*$_{\text{-masc}}$, *die*$_{\text{-fem}}$, *das*$_{\textit{-neut}}$, and the indefinite articles *ein*$_{\text{-masc}}$, *eine*$_{\text{-fem}}$, and *ein*$_{\textit{-neut}}$. In German, definite and indefinite determiners as well as demonstratives, personal pronouns, relative pronouns, and attributive adjectives have to agree with the noun´s gender (and number). Predicative adjectives are not specific in gender (and number). Contrary to the Spanish gender system, a noun´s gender can be difficult to infer from its morpho-phonological form or semantics. Sometimes, however, morpho-phonological and/or semantic cues are available: “… there is a complex interplay of overlapping semantic, morphological and phonological factors.” (Corbett, 1991, p. 49). Two of the most well-known rules[7], for example, are the phonological rule that nouns ending in /ə/ (schwa) are usually, but not always, feminine, and the morphological rule that nouns ending in *-er* (a derivational suffix) are usually, but not always, masculine. Some more morphological rules (for polysyllabic nouns) are summarized in German grammars: typically feminine derivational suffixes that are mentioned are *-ei*, *-heit*, *-keit*, *-schaft, -ung*, typically masculine derivational suffixes are *-ist*, *- ismus*, *-ler*, *-ling*, *-rich*, *-or*, and typically neuter

---

[3] Exceptions are, for example, *el aroma the*$_{\textit{-masc}}$ *aroma, el mapa–the*$_{\textit{-masc}}$ *map, el planeta–the*$_{\textit{-masc}}$ *planet, la mano–the*$_{\textit{-fem}}$ *hand, la dynamo–the*$_{\textit{-fem}}$ *dynamo.*

[4] Even though there are also many feminine nouns ending in *-e* (Vera Morales, 2013, p. 15): *la nube–the*$_{\textit{-fem}}$ *cloud*, *la fiebre–the*$_{\textit{-fem}}$ *fever*, *la leche–the*$_{\textit{-fem}}$ *milk*, *la fe–the*$_{\textit{-fem}}$ *faith*, *la calle–the*$_{\textit{-fem}}$ *street*, *la carne–the*$_{\textit{-fem}}$ *meat*, *la gripe–the*$_{\textit{-fem}}$ *flu*, *la suerte–the*$_{\textit{-fem}}$ *luck*. It is not clear if these cases all constitute exceptions or how they can be reconciled with Bergen´s rule. The other morpho-phonological rules are in agreement with Vera Morales´ summary of Spanish gender rules.

[5] For example, *la mosca–the*$_{\textit{-fem}}$ *fly*, *el caballo–the*$_{\textit{-masc}}$ *horse*.

[6] Furthermore, as discussed in section 3.3.1, learning any L2 gender system, however simple it may be, seems to be more difficult for native speakers of an ungendered language than for speakers of a gendered language.

[7] The rule that grammatical gender usually coincides with sex and the morphological rule that nouns ending in *-chen* (diminutives) are neutral (which overrides the former rule) are two other well-known rules.

gender derivational suffixes *-chen*, *-lein*, *-ment*, *- nis*, *-tum*, *-um* (Eisenberg, 2013; Götze & Hess-Lüttich, 1999; Hoberg & Hoberg, 2009). Examples for semantic rules include the circumstance that most names of minerals and types of rock (*der Fels*, *der Flint*, der *Kalk*) as well as most alcoholic drinks (*der Rum*, *der Gin*) are masculine, while basic numbers are feminine (*die Eins*, *die Zwei*), and physical and theoretical units are neutral (*das Lux*, *das Ohm*) as well as chemical materials and metals (*das Blei*, *das Sulfat*) (Köpcke & Zubin, 1983). So it seems that even if the German gender system is not completely arbitrary (as it was previously stated by some researchers, cf. e.g., Bloomfield (1933) and Brinkmann (1962)), the rules seem to be much more complex than, for example, in Spanish and include many exceptions for which no rules can be defined.

This is further illustrated by the results of a study conducted by Köpcke (1982). Köpcke analyzed the rules governing gender assignment of the 1466 nouns with one syllable that appeared in the "Duden" dictionary and was able to explain about 90 % of the gender assignments with 24 phonological, 5 (plural-) morphological, and 15 semantic rules. The implicit knowledge of these rules in native speakers was later experimentally verified with nonwords by Köpcke and Zubin (1983). This shows that there might be rules governing the gender assignment in German, but that these rules are disproportionately complex in comparison to Spanish: For the gender assignment of 90 % of monosyllabic German nouns, a greater set of rules has to be proposed than for 97.3 % of investigated Spanish nouns examined by Bergen (1978). Not surprisingly, the acquisition of the German gender system by L2 learners is usually viewed as difficult (Rogers, 1987; Spinner & Juffs, 2008).

What complicates the acquisition of the German gender system even more is its complex declensional system. Form to meaning mapping for nouns and their genders is handicapped by the fact that, as Rogers (1987) points out, gender, number, and case markings in determiners are conflated. The reason for this is that only six determiners forms (*die, der, das, dem, des*, and *den*) are available to express sixteen different combinations of number, gender, and case (p. 50). The form *der*, for example, is used as a determiner for nominative masculine nouns, but it is also assigned to feminine nouns in genitive and dative singular and to all genders in genitive plural. According to Taraban, McDonald, and MacWhinney (1989), this many to many cue-to-category mapping leads to low cue strength, that is, the determiner *der* has a very low informational value for the L2 learner as to what gender it refers to. Furthermore, the declension of determiners might lead to a low input frequency of the unambiguously gender-marked nominative article[8]. As discussed in section 1.5.2 on the Competition Model, reliable and frequent cues are more valid and stronger than ambiguous and infrequent cues. Thus, the German gender system provides only weak cues of low validity for the L2 learner. Therefore, the Competition Model would predict that the Spanish gender system should be easier to acquire than the German gender system. Also, DeKeyser (2005) identified complexity of form–meaning relationships as one of the most important factors that make learning L2 grammar difficult. Nevertheless, because of the difficulties associated with the acquisition of the German gender system, it can be especially interesting for L2 acquisition research, especially in comparison with a fairly unproblematic gender system, such as the Spanish system. For this reason, and for other reasons mentioned before, Spanish and German were chosen for the first experiment. In the following, the predictions for the first experiment will be specified.

[8] Singular nominative determiners are unambiguously gender-marked except for the masculine and neuter indefinite article which is *ein* for both genders and therefore gender-ambiguous.

4.1.2 Predictions

As discussed in section 3.3.2, gender interference in language production is usually investigated in one of two syntactic structures: NP or bare noun production. It appears that for different languages, a different operationalization of gender interference is suitable. Those Romance languages with transparent gender systems show the monolingual gender interference effect in bare noun production, Germanic languages show this effect in NP production. Consequently, for the investigation of gender interference between a Romance and a Germanic language, it seems that both structures would have to be investigated. As also discussed in section 3.3.2, interference in NPs is usually investigated in comparison to a bare noun condition (Bordag & Pechmann, 2007; Lemhöfer, Schriefers, & Hanique, 2010; Lemhöfer, Spalek, & Schriefers, 2008). Here, an inhibitory priming effect was expected for the incongruent conditions, compared to the congruent condition, and relative to the bare noun condition. Interference in bare noun naming, on the other hand, can only be investigated in comparison to the performance of a monolingual control group (cf. Paolieri, Cubelli, et al., 2010). Here, also an inhibitory effect was expected for the incongruent conditions[9], relative to the monolingual control group.

Furthermore, besides the RTs, gender transfer effects should also be visible in the error rates, with more errors in the incongruent conditions than in the congruent conditions (cf. Lemhöfer et al., 2010; Lemhöfer et al., 2008). It is also expected that transfer effects will be stronger at lower proficiency levels than at higher proficiency levels. Moreover, it is likely that stronger transfer effects will be observed for cognates than for noncognates. Regarding the direction of gender transfer – that is, whether transfer will occur only from L1 German into L2 Spanish, vice versa, or between both languages – no specific predictions were made.

As mentioned before, Experiment 1 consists of three parts: a PNT, an offline gender assignment task, and an LDT. In the following, first the General Method and the Method, Results, and Discussion of the PNT will be described. After that, the Method, Results, and Discussion of the offline gender assignment task will be explained. Finally, some theoretical background regarding the LDT will be discussed and the Method, Results, and Discussion of the LDT will be described.

[9] Note, however, that since this expectation is based on the results of the PNT of Paolieri, Cubelli, et al. (2010), the effect could be different for the LDT.

4.2 General method and method PNT

4.2.1 Participants

In total, 174 participants took part in the study:

- 47 native German speakers who spoke Spanish as an L2 in the Spanish version of the experiment
- A control group of 40 monolingual native speakers of Spanish, 20 in each (online) task (LDT/PNT)
- 47 native Spanish speakers who spoke German as an L2 in the German version of the experiment
- A control group of 41 monolingual native speakers of German, 20/21 in each (online) task (LDT/PNT)

The LDT and the PNT were administered in two different sessions. The bilingual[10] speakers participated in both sessions. The monolingual speakers participated in only one of the online tasks. All participants had normal or corrected-to-normal vision. All participants were either tested at the Humboldt University of Berlin in Germany or at the University of Murcia in Spain.

The bilingual speakers received € 10 as a reimbursement after participating in the second session. The monolingual German speakers received € 5 as they only participated in one session each. The monolingual Spanish speakers were tested at the Universidad de Murcia in Spain and received course credit for participation.

Exclusion criteria:

L2 proficiency was measured at the beginning of each experimental session with the grammar test of the DIALANG test (cf. section 4.2.3). The minimum L2 competence required for participation in this study was A2 according to the Common European Frame of Reference[11] (CEFR; Council of Europe, 2001) and a score of at least 150 in the vocabulary placement test of the DIALANG test (Alderson & Huhta, 2005; Alderson, 2006).

Only bilingual subjects who participated in both tasks (LDT and PNT) were included in the analysis because only then all the metadata (cf. section 4.2.3) would be available. However, if the RT data of one of the tasks could not be used or was unavailable due to technical problems, the data of the subject could still be included in the analysis of the other task since all metadata would be available.

In the following section, the participants will be described in more detail.

[10] "Bilingual" here means "late bilingual", unless specified otherwise (e.g., "early bilingual").
[11] The CEFR scale ranges from the lowest proficiency level A1 (beginner) to the highest proficiency level of C2 (native-like). Levels are A1, A2, B1, B2, C1, and C2.

Monolingual control groups

Monolingual Spanish speakers

In the LDT, 19 subjects were female, 1 male. In the PNT, 18 subjects were female, 2 male. All participants were psychology students currently enrolled at the University of Murcia, Spain. Their age was not collected but ranged approximately from 20 to 25 years.

Monolingual German speakers

In the monolingual control groups, each task (LDT and PNT) was carried out separately. 20 subjects participated in the LDT and 21 subjects in the PNT. Since the data of one subject had to be excluded because she had received the wrong prime list, 20 participants entered data analysis in the PNT. In the LDT, 11 subjects were female, 9 male. Their age ranged from 20 to 46, mean age was 29.3. In the PNT, 13 subjects were female, 7 male. Their age ranged from 20 to 42, mean age was 27.9. Participants were invited through the subject database of the Humboldt University of Berlin.

Bilingual experimental groups

German bilinguals

47 native German speakers, all born and raised in Germany[12], were tested. Three subjects had to be excluded: Two subjects´ language competence was too low (lower than A2 according to the CEFR), one subject did not participate in the second session.

In the data analysis of the LDT, no more participants had to be excluded. The analysis of the LDT was thus based on 44 subjects. The analysis of the bilingual with the monolingual group together (bare noun effect) was conducted with the same 44 participants.

In the analysis of the PNT, four more subjects had to be excluded: Of one subject there was no data available due to a technical error and three subjects had to be excluded due to missing observations in some cells after error removal. 40 subjects entered the final data analysis. The analysis of the bilingual with the monolingual group together was based on the same participants. The error analysis was based on 43 subjects because no subjects had to be excluded due to missing observations.

The following description of the subjects is based on the 44 subjects who were included in the data analysis of the LDT:

35 of the subjects were female (80 %), 9 male. Two subjects were left-handed. One subject stated that German and Spanish were her mother tongues and that she had learned Spanish at age 0. Due to her low language competence (B1), however, she was not considered a genuine early bilingual. Eight subjects were tested in Spain (Universidad de Murcia), the rest were tested at the Humboldt University of Berlin (psycholinguistic laboratory). Quite a few of the subjects also had some knowledge of other Romance languages because French and/or Latin are usually obligatory languages in High School in Germany. This could be a problem because interference in L2 processing can also arise from other foreign languages spoken. However, generally, Spanish was the most dominant language of the Romance languages so that interference effects should be rather small[13].

[12] One subject stated that she was also raised in Ethiopia till age 8.

[13] Except for four cases where Portuguese was used as often as or more often than Spanish. Nevertheless, examination of Portuguese translation equivalents revealed that only nine Portuguese nouns had a different

42 subjects had lived in a Spanish-speaking country (2 months or more[14]). All of the subjects were either students currently enrolled at a university or had studied at a university before so that their level of education was similar. The time span between the two experimental tasks was at least 7 days and maximum 55 days, mean distance was 15 days.

	Mean	SD	Min	Max
Age	28.02	6	21	51
Age of acquisition of Spanish	18.02	5.16	0	27
Months spent in Spanish-speaking country	35.57	67.01	0	300
Number of foreign languages spoken at the time of starting to learn Spanish	2.50	1.13	0	5
Number of foreign languages spoken at the time of testing (including Spanish)	3	0.91	2	5
Reading frequency in Spanish	5.64	2.98	0	10
Speaking frequency in Spanish	6.43	2.71	1	10
Frequency of watching Spanish TV or listening to Spanish radio	3.39	3.66	0	10
Frequency of writing Spanish	5.64	2.85	0	10
Spanish reading proficiency	7.43	1.98	3	10
Spanish-speaking proficiency	7.27	1.77	4	10
Proficiency of understanding spoken Spanish	7.89	1.73	3	10
Spanish writing proficiency	6.73	2.14	3	10

Table 4.1 Results of the language history questionnaire of bilingual German subjects. Frequency and proficiency ratings were indicated on scales ranging from 0 (not at all/very low) to 10 (very often/very high).

An overview of the self-rated language competence and other metadata of the 44 subjects who entered data analysis of the LDT is given in Table 4.1.

gender from Spanish nouns. Therefore, if these subjects experienced gender interference from Portuguese, this effect should be rather small.

[14] The amount of time spent in a Spanish-speaking country is shorter for German subjects than the amount of time Spanish subjects spent in a German-speaking country because Spanish subjects were mostly tested in Germany where they had already been living for some time. However, quite a few of the German subjects had also spent half a year or a year in a Spanish-speaking country (cf. Table 4.1).

Proficiency	Level CEFR	Number of subjects	Subjects per Level
“Low”	A2	7	23
	B1	16	
“High”	B2	16	21
	C1	4	
	C2	1	

Table 4.2 Language competence of the subjects according to the CEFR levels and number of subjects per Level (low- or high-proficient).

Bilinguals who achieved at least level B2 on the grammar test were considered as high-proficient and bilinguals with level B1 or lower (at least A2) were considered as low-proficient. The difference between the two Levels (low- and high-proficient) was significant regarding the CEFR levels (A2 - C2) ($X^2 = 44.000$, df = 4, $p < 0.001$).

Spanish bilinguals

47 native Spanish speakers, from Spain and Latin America[15], were tested. Seven subjects had to be excluded before data analysis: Six subjects did not participate in the second session so that some of the metadata could not be collected and one subject´s language competence was too low.

In the data analysis of the LDT, another participant had to be excluded due to missing observations in some cells after error removal. The analysis of the LDT was thus based on 39 subjects. The analysis of the bilingual with the monolingual group together was based on the same 39 subjects.

In the analysis of the PNT, of the 39 subjects, one more participant had to be excluded due to technical problems with the recording. The error analysis was thus based on 38 subjects. Regarding the RT analysis, the bilingual analysis investigating the NP effect was not conducted due to certain effects in the monolingual analysis[16]. The NP effect was instead investigated in an analysis with the monolingual and bilingual group together, just as the bare noun effect (cf. section 4.2.4). In these analyses, 8 more subjects had to be excluded due to missing observations so that these analyses were conducted with 30 bilingual subjects each.

The following description of the subjects is based on the 39 subjects who were included in the data analysis of the LDT:

18 (46 %) of the subjects were female, 21 subjects were male. Three participants were left-handed. All participants´ mother tongue was Spanish. One subject stated that Spanish and Galician (a dialect spoken in the region of Galicia in northwestern Spain) were her native languages. Four subjects were tested in Spain (laboratory of the Cognitive Science research group at the Universidad de Murcia), the others were tested at the Humboldt University of Berlin (psycholinguistic laboratory). All subjects were either currently living in Germany (for at least 6 months) or had lived in a German-speaking country before for at least 1.5 years. Fourteen reported to have been born and raised in Spain, one was raised in Spain and Germany, the remaining 24 were from Latin American countries. Some

[15] For the sake of brevity, I might refer to the Spanish-speaking participants as “Spanish subjects“, “Spanish bilinguals“, or “Spaniards” which also means to include the Spanish-speaking participants from Latin America.

[16] The reason for this will be explained in section 4.3.2.

subjects also had some knowledge of other Romance languages but Spanish was clearly their most dominant language. As all of the subjects (two not specified) either were students currently enrolled at a university or had studied at a university before, their level of education was similar. The time span between the two experimental tasks was at least 6 days and maximum 40 days, mean distance was 14 days.

	Mean	SD	Min	Max
Age	28.64	6.04	20	52
Age of acquisition of German	20.13	6.73	4	32
Months spent in German-speaking country	48.01	35.73	6	156
Number of foreign languages spoken at the time of starting to learn German	1.90	0.72	1	3
Number of foreign languages spoken at the time of testing (including German)	2.82	1.04	1	5
Reading frequency in German	8.05	2.47	1	10
Speaking frequency in German	5.97	3.40	0	10
Frequency of watching German TV or listening to German radio	6.63	2.83	1	10
Frequency of writing German	7.63	1.73	5	10
German reading proficiency	7.45	1.64	5	10
German-speaking proficiency	8.45	1.43	5	10
Proficiency of understanding spoken German	6.51	1.80	2	10
German writing proficiency	8.05	2.47	1	10

Table 4.3 Results of the language history questionnaire of bilingual Spanish subjects. Frequency and proficiency ratings were indicated on scales ranging from 0 (not at all/very low) to 10 (very often/very high).

An overview of the self-rated language competence and other metadata of the 39 subjects who entered data analysis of the LDT is given in Table 4.3.

An overview of the Spanish bilinguals´ proficiency scores as measured in the DIALANG test is given in Table 4.4.

Proficiency	Level CEFR	Number of subjects	Subjects per Level
“Low”	A2	3	19
	B1	16	
“High”	B2	18	20
	C1	0	
	C2	2	

Table 4.4 Language competence of the subjects according to the CEFR levels and number of subjects per Level (low- or high-proficient).

As explained above, bilinguals who achieved at least level B2 on the grammar test were considered as high-proficient and bilinguals with level B1 or lower were considered as low-proficient. The difference between the two Levels (low- and high-proficient) was significant regarding the CEFR levels ($\chi^2 = 39.000$, df = 3, $p < 0.001$). Furthermore, distribution of language proficiency (CEFR levels A2 - C2) in the two proficiency groups was similar across German and Spanish subjects ($\chi^2 = 5.771$, df = 4, $p = 0.217$).

4.2.2 Material

The experimental material consisted of 114 concrete nouns (72 noncognates, 42 cognates) that could be easily depicted and belonged to the basic vocabulary of an L2 learner (cf. next paragraph). A complete list of experimental stimuli can be found in section 10.1. The same nouns in Spanish and in German were used. There were three “Gender Compatibility conditions”: one gender-congruent condition and two gender-incongruent conditions. In the congruent condition, nouns had the same grammatical gender in German and Spanish. In the two gender-incongruent conditions, nouns had different grammatical genders across the two languages. In one of the two incongruent conditions, termed the “incongruent condition”, nouns with female gender in German had masculine gender in Spanish and vice versa. In the second incongruent condition, the “incongruent neutral condition”, nouns had neutral gender in German but masculine or female gender in Spanish. For noncognates, there were 24 nouns in each Gender Compatibility condition, for cognates 14 nouns. Items were judged as cognates when they were orthographically and/or phonologically very similar in the two languages, while for noncognates this similarity was small. In noncognate conditions, the number of feminine and masculine items was balanced. For cognates, there were five feminine and nine masculine Spanish items in the incongruent condition, two feminine and twelve masculine Spanish items in the incongruent neutral condition. Due to the limited number of incongruent cognates that would belong to the basic vocabulary of a language learner and that could be depicted, this imbalance could not be avoided.

In order to assure that even beginning L2 learners would know most of the items, I tried to mainly select nouns from basic vocabulary. The learner dictionary “Langenscheidt Grundwortschatz Spanisch” (Langenscheidt Basic Vocabulary Spanish, Duenas de Haensch, 1999) was used as a reference as it comprises a vocabulary selection of 4000 words for beginning learners of Spanish, selected according to “frequency, up-to-dateness and practical value” (p. 9). Yet, because of the many requirements the material had to fulfill, it was not possible to only use words from the Langenscheidt Grundwortschatz. Of the final selection, 90 % of the items in all the noncognate conditions and 62 % of the items in the cognate conditions belonged to Langenscheidt´s basic vocabulary selection. Furthermore, as mentioned above, only concrete nouns were used because they can usually be depicted well. Research has also shown that concrete words are easier to learn (Groot & van Hell, 2005), which makes it more likely that they are known even by low-proficient language speakers.

Items were matched across both languages and across conditions for number of letters and syllables and (logarithmic) item frequency. Frequency measures were obtained from the CELEX database (Baayen, Piepenbrock, & Gulikers, 1995). Because the CELEX frequency data did not seem to intuitively match the experience of an L2 learner[17], I decided to additionally use the frequency data[18] of Projekt Deutscher Wortschatz (PDW) of the University of Leipzig (e.g., Biemann, Bordag, Heyer & Quasthoff, 2004; http://wortschatz.uni-leipzig.de/). For the Spanish items, frequency data were obtained from "BuscaPalabras" (Davis & Perea, 2005; http://www.uv.es/mperea/)[19]. A matching across cognate status, however, could not be achieved as cognate words tended to have more letters and syllables than noncognates (cf. Table 4.5 and Table 4.7). A complete list of the items can be found in Appendix 10.1. Spanish items were also matched regarding gender transparency of the noun endings[20]. However, due to the great number of matching criteria, for German items, ending transparency (e.g., /ə/ (schwa), cf. section 4.1.1) for a fairly salient transparent ending) could not be considered in the matching process[21]. Matching was achieved by using the software Match (Van Casteren & Davis, 2007).

[17] Quite a few very common words that – in my opinion – should be known by even beginning L2 learners such as *Rücken* (back) or *CD* have a frequency of 0 (i.e., do not appear) in the CELEX data base, while a few very uncommon words have a higher frequency, e.g., *Rückblick* (retrospection, frequency = 0.77820) or *Schädelbruch* (skull fracture, frequency = 0.30100). The frequency measures of PDW did at least in some cases seem more convincing in comparison: *Rücken* (7634, frequency class: 9), *CD* (2855, frequency class: 11), *Rückblick* (945, frequency class: 12), *Schädelbruch* (77, frequency class: 16). In each case, none of the two corpora were developed to reflect word exposure of language learners but with using two different corpora for the German words (in addition to the Spanish corpus) I tried to obtain more representative frequency measures. Therefore, all items were matched for three different frequency measures.

[18] In the PDW corpus, absolute frequency counts of the query word in the corpus are given, e.g., *Haus* (house) has a frequency of 32011 in the corpus. In addition, a frequency class is reported. The frequency class of *Haus* is 7, which means that *der* (the most frequent word in the German language) is approximately 2^7 times more frequent than *Haus*. Experimental items were matched using the logarithmic frequency and the frequency class of words.

[19] The frequency data of BuscaPalabras, in turn, are based on the frequency data of the LEXESP corpus (Sebastián-Gallés, Martí, Cuetos, & Carreiras, 2000).

[20] The simplest and probably first rule a beginning L2 learner of Spanish learns regarding noun endings and grammatical gender in Spanish is that nouns ending in *-o* are usually masculine and nouns ending in *-a* are usually feminine. Even though there are more rules regarding noun endings and grammatical gender, only these two most salient gender-marked noun endings were taken into account for matching transparency across conditions.

[21] Note that after matching was completed, unfortunately, it turned out that overall (for pooled noncognates and cognates) there were more German nouns with the transparent ending /ə/ in the congruent condition (14; 53.8 %) than in the incongruent (12; 46.2 %) and incongruent neutral condition (0; 0 %). This difference across conditions was significant ($F(2, 111) = 9.821$, $p < 0.001$) presumably rendering the congruent condition easier than the incongruent conditions. However, *t*-tests (Bonferroni-corrected) revealed that the difference between the congruent and incongruent condition was not significant ($t = 0.478$, df = 74, $p = 1.000$). Only the differences between the congruent and incongruent neutral condition and the incongruent and incongruent neutral condition was significant ($t = 4.646$, df = 74, $p < 0.001$and $t = 4.132$, df = 74, $p < 0.001$, respectively). Considering that the error rates in the offline gender assignment task (cf. section 4.5.2) were 15 % in the congruent

In the following section, an overview of the characteristics of the noncognate and cognate items in the different conditions is given.

Noncognates

An overview of the characteristics of the Spanish and German noncognate items is given in Table 4.5.

		Congruent		Incongruent		Incongruent n	
		Mean (SD)	**Range**	**Mean (SD)**	**Range**	**Mean (SD)**	**Range**
Spanish	**Frequency BuscaPalabras**	1.28 (0.47)	0.19 - 2.43	1.30 (0.60)	0.28 - 2.83	1.40 (0.45)	0.47 - 2.29
	# of syllables	2.54 (0.59)	1 - 3	2.50 (0.59)	2 - 4	2.67 (0.70)	1 - 4
	# of letters	5.88 (1.60)	3 - 9	6.04 (1.30)	4 - 8	6.25 (1.57)	3 - 9
German	**Frequency CELEX**	1.39 (0.54)	0.30 - 2.61	1.25 (0.75)	0 - 2.99	1.39 (0.40)	0.78 - 2.26
	Frequency PDW	3.19 (0.61)	2.18 - 4.84	3.20 (0.74)	1.89 - 4.98	3.31 (0.41)	2.61 - 4.19
	Frequency class PDW	11.58 (2.04)	6 - 15	11.63 (2.41)	6 - 16	11.17 (1.34)	8 - 13
	# of syllables	1.58 (0.50)	1 - 2	1.58 (0.50)	1 - 2	1.38 (0.50)	1 - 2
	# of letters	5.17 (1.13)	3 - 8	4.92 (1.44)	2 - 8	5.21 (1.35)	3 - 8

Table 4.5 The mean, standard deviation, and range of the frequency data of the different corpora, number of syllables and letters of Spanish and German noncognates per Gender Compatibility condition (congruent, incongruent, incongruent neutral). All frequencies are logarithmic frequencies.

Of the Spanish items, 83 % (20 out of 24) had a transparent gender ending in the congruent condition and 88 % (21 out of 24) in each of the incongruent conditions.

An ANOVA was carried out with the variables number of syllables, number of letters, and the (logarithmic) frequency measures of the three corpora (BuscaPalabras, CELEX, PDW). There were no significant differences between the conditions regarding any of these variables (all $F < 1$, except for number of syllables in German ($F = 1.386$), where the incongruent neutral condition had a lower number of syllables on average, cf. Table 4.5). The results of the ANOVA are summarized in Table 4.6.

condition, 25 % in the incongruent condition, and 28 % in the incongruent neutral condition, this imbalance across conditions probably had little biasing effect on the results.

		df	*F*	*p*
Spanish	**Transparency**	2,69	.112	.895
	Frequency BuscaPalabras	2,69	.347	.708
	# of syllables	2,69	.456	.635
	# of letters	2,69	.379	.686
German	**Frequency CELEX**	2,69	.496	.611
	Frequency PDW	2,69	.273	.762
	Frequency class PDW	2,69	.393	.677
	# of syllables	2,69	1.386	.257
	# of letters	2,69	.346	.709

Table 4.6 Results of the ANOVA for logarithmic frequencies of the different corpora, number of syllables and letters of Spanish and German noncognates, as well as for transparency of noun endings for Spanish items.

Cognates

An overview of the characteristics of the Spanish and German cognate items is given in Table 4.7.

		Congruent		**Incongruent**		**Incongruent n**	
		Mean (SD)	**Range**	**Mean (SD)**	**Range**	**Mean (SD)**	**Range**
Spanish	**Frequency BuscaPalabras**	1.04 (0.39)	0.19 - 1.61	1.06 (0.54)	0.19 - 2.27	1.13 (0.37)	0.44 - 1.81
	# of syllables	2.79 (0.80)	2 - 4	3.07 (1.14)	1 - 5	2.71 (0.83)	1 - 4
	# of letters	6.29 (1.59)	4 - 8	6.86 (2.45)	3 - 11	6.36 (2.17)	3 - 11

German	**Frequency CELEX**	0.95 (0.42)	0.48 - 1.97	0.97 (0.63)	0 - 2	1.01 (0.49)	0 - 2.07
	Frequency PDW	3.02 (0.41)	2.44 - 3.77	3.00 (0.89)	1.11 - 4.54	3.04 (0.62)	2.21 - 4.02
	Frequency class PDW	12.21 (1.37)	10 - 14	12.07 (2.87)	7 - 18	12.07 (2.06)	9 - 15
	# of syllables	6.29 (1.33)	4 - 8	6.50 (1.91)	4 - 10	5.86 (1.79)	4 - 9
	# of letters	2.57 (0.76)	1 - 4	2.71 (0.99)	1 - 4	2.36 (0.63)	1 - 3

Table 4.7 The mean, standard deviation, and range of the frequency data of the different corpora, number of syllables and letters of Spanish and German cognates per Gender Compatibility condition (congruent, incongruent, incongruent neutral). All frequencies are logarithmic frequencies.

Of the Spanish items, in each condition 43 % (6 out of 14) of the items had a transparent gender ending.

An ANOVA revealed that there were no significant differences between the conditions regarding any of these variables (all F < 1). The results of the ANOVA are summarized in Table 4.8.

		df	*F*	*p*
Spanish	**Transparency**	2,39	.000	1.000
	Frequency BuscaPalabras	2,39	.176	.839
	# of syllables	2,39	.571	.570
	# of letters	2,39	.308	.737
German	**Frequency CELEX**	2,39	.048	.953
	Frequency PDW	2,39	.012	.988
	Frequency class PDW	2,39	.020	.980
	# of syllables	2,39	.522	.597
	# of letters	2,39	.692	.507

Table 4.8 Results of the ANOVA for logarithmic frequencies of the different corpora, number of syllables and letters of Spanish and German cognates, as well as for transparency of noun endings for Spanish items.

The conditions were very similar also according to more subjective measures. In a translation task (cf. section 4.2.3), more than 85 % of the Spanish and German items in each condition were translated correctly and with the intended translation equivalent. Word familiarity, as indicated in the offline gender assignment task (cf. section 4.5.1), was also very high and similar across conditions. More than 85 % of the words received a score of no less than 6 out of a 0 to 7 scale (0 =

not familiar at all, 7 = very familiar) in every condition for the German items. For the Spanish items, more than 85 % of the words received a score of at least 5 in every condition, except for the cognate incongruent neutral condition. In this condition, items were apparently somewhat less familiar and only about 75 % received a score of at least 5 and approximately 85 % received a score of at least 4.

Material PNT

The material for the PNT consisted of black and white line drawings of objects depicting the nouns also used in the LDT. Pictures were obtained from different sources: from the online picture data base of the International Picture-Naming Project at the Center for Research in Language of the University of California in San Diego (Szekely et al., 2004; available for download at: http://crl.ucsd.edu/experiments/ipnp/method/browsepics/separately/view.html), with kind permission from Xavier Alario (Alario, Ferrand, et al., 2004) and from Katharina Spalek (Lemhöfer, Spalek, & Schriefers, 2008). Additional pictures were obtained from the internet and changed with picture editing software to match the other pictures. Picture size was approximately 6 x 6 cm (200 - 250 x 200 - 250 pixels) with a resolution of 37.8 pixels per cm. The pictures were presented as black line drawings on a white background. As practice items, 24 additional pictures were used.

In total, there were 114 pictures, 64 with Spanish masculine gender, 50 with Spanish feminine gender. In German the distribution was as follows: 36 with masculine gender, 40 with feminine gender, 38 with neutral gender. There were 72 noncognates and 42 cognates. Additionally, at the beginning of both experimental blocks, two warming-up pictures which were not analyzed were presented. (Two of the warming-up pictures had German masculine gender, two had German neutral gender. All four had Spanish feminine gender.

4.2.3 General procedure

For the bilingual subjects, in addition to the experimental tasks, the PNT, LDT, and offline gender assignment task, an experimental session consisted of various other parts in order to collect additional metadata: the DIALANG language proficiency test, a word translation task, and a language history questionnaire. The order of tasks for the bilingual subjects is depicted in Figure 4.1. The DIALANG test was always conducted prior to the experiment and will be described below. The offline gender assignment task was always conducted after the PNT and will be described in section 4.5. The word translation task and language history questionnaire were administered after the LDT and will be described below.

Prior to participation, subjects signed an informed consent form, and the bilingual subjects completed the DIALANG test (Alderson & Huhta, 2005; Alderson, 2006; see below) in order to assess their language competence. Subsequently, either the LDT or the PNT were conducted. Of the experimental groups, half of the subjects participated first in the LDT and then in the PNT and the other half first in the PNT and then in the LDT, so that the order of tasks was counterbalanced across subjects. The monolingual control groups participated in only one of the tasks and did not complete the DIALANG test nor any of the other additional offline tasks. Two prime lists were created. Subjects assigned to one of the prime lists named or read (according to the task, PNT or LDT, respectively) one half of the nouns per condition with a determiner (NP processing) and the other half without a determiner (bare noun processing). For subjects who received the complementary prime list, the assignment was reversed. In the second session, that is, for either the PNT or the LDT, subjects received the prime list complementary to the one of the first session.

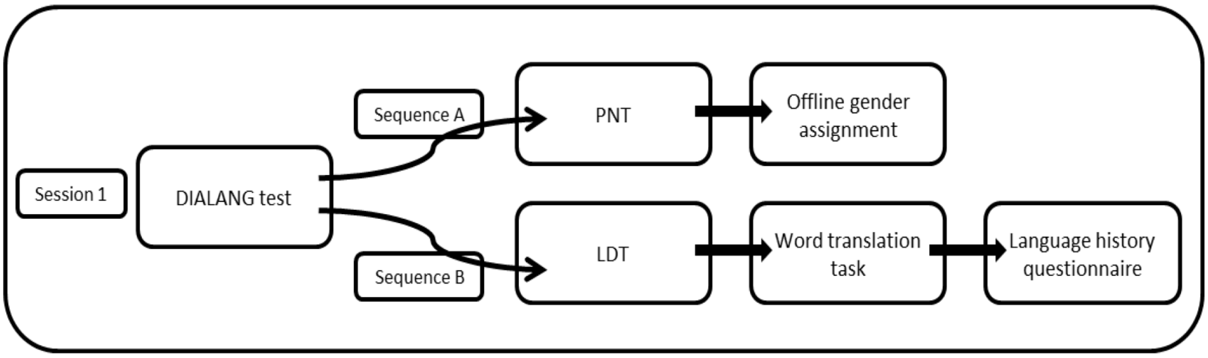


Figure 4.1 The first session always started with the DIALANG language test. Then, participants followed either sequence A or B. In the second session, participants completed the sequence they had not completed before.

Experimental items were pseudorandomized. Pseudorandomization was achieved using the Software Mix (van Casteren & Davis, 2006). The following randomization constraints were employed (across both the German and the Spanish version of the experiment): There was a minimum distance of three words between words starting with the same letter. Maximum repetition of the same Phrase Type (PNT) or Prime Type (LDT), that is, bare noun vs. NP, was limited to four. The same Spanish article could appear four times in a row, the same German article five times in a row (both randomization criteria were used for both language versions). Noncognates could appear five times in a row, cognates a maximum of two times. In the LDT, the same word type (word/nonword) was repeated at most five times. Each participant received a different pseudorandomized version of the stimuli. The same pseudo-randomization criteria were used for both LDT and PNT, except that there were no nonwords in the PNT.

The participants were tested individually in a quiet room, seated on a chair. The visual stimuli were presented centered on a Belinea 17-inch LCD-monitor with a data resolution of 1024 by 768 pixels. Viewing distance was about 50 cm. Words and pictures were presented with the experimental software DMDX (Forster & Forster, 2003), version 3.3.0.2 running on an IMTEC computer with an Intel(R) Pentium(R) 4 CPU 3.06GHz (2 CPUs) processor and Windows XP Professional operating system. The display device was an ATI RADEON 7000 AGP. In the LDT a Cherry keyboard, model RS 6000 M, was used for response registration. For the PNT, a Sennheiser Headset PC131 was used.

Experimental sessions without the DIALANG test lasted 30 - 45 minutes. If the session was the first session, and therefore the DIALANG test was administered, the session took about 1 to 1.5 hours in total.

Language competence – DIALANG test

The DIALANG test was always carried out prior to the experiment. The main reason for the decision to use the DIALANG test was that the different ways of assessing L2 proficiency employed in previous studies oftentimes render comparisons between studies difficult. In the hope of making the proficiency scores of the present study more comparable and interpretable, I sought to use a test based on the Common European Framework of Reference (CEFR; Council of Europe, 2001). The CEFR is used EU-wide to specify level of language proficiency as well as for course placement in many language schools and could therefore provide a way of standardizing the measurement of L2 proficiency across experiments. The DIALANG test (available at: http://www.lancs.ac.uk/researchenterprise/DIALANG/about) has been developed by many European higher education institutions with the support of the European Commission and test results are reported on the six levels of the CEFR scale (A1, A2, B1, B2, C1, C2) with A1 being the lowest level (beginner) and C2 the highest level (native-like). In the present experiment, bilinguals who achieved at least level B2 on the grammar test were considered high-proficient and bilinguals with level B1 or lower (at least A2) were considered low-proficient.

DIALANG is an on-line language assessment system which contains tests in 14 European languages. The five competence areas of reading, writing, listening, grammar, and vocabulary can be tested in each of the languages. (For more information about the validity, reliability and calibration of the DIALANG tests, cf. Alderson & Huhta, 2005; Alderson, 2006). I decided to use the grammar test as I thought this test would best match the focus of the experiment, that is, the mastery of grammatical gender. The grammar test contained 30 tasks or questions, where usually the correct grammatical form of, for example, verbs and articles had to be filled in. Prior to each test, the DIALANG test presents a short placement test consisting of a vocabulary test of 75 items (verbs) in which existing words in the target language have to be distinguished from nonwords. Subsequently, according to their test score (ranging from 0 to 999), participants receive one of three versions of the grammar test varying in difficulty. There was no time limit, but on average, it took participants about 45 minutes to complete both tests.

Word translation task

The word translation task was conducted after the LDT. In this task, participants had to translate all the experimental items into their native language in order to make sure that they were familiar with all the words. Words that were unknown, translated wrongly, or translated with a different translation than the intended translation were removed from this participant´s analysis.

For clarification: Provided translation equivalents that did not appear as possible translation in the online dictionary www.leo.org were considered "wrong translations" (such as e.g., *montaña* (*mountain*)–*Hügel* (*hill*)). Translation equivalents that were listed in the online dictionary but diverged from the expected translation for the purpose of the experiment were considered "different translations". These "different" translations included cases where a synonym of the expected translation equivalent was provided. For example, *Berg (mountain)* can be translated with either *monte*$_{\text{-masc}}$ or *montaña*$_{\text{-fem.}}$ Sometimes these different translations were due to regional variations, such as the word *Avocado* (avocado) which is *aguacate*$_{\text{-masc}}$ in Spain and Mexico but *palta*$_{\text{-fem}}$ in Argentina, Chile, Peru, and Uruguay. Excluding these different translations was especially important when they differed in gender, such as in the examples given here. Items where no translation equivalent was provided, that is, where the slot was left empty, were labeled "unknown translations".

Language history questionnaire

After completing the word translation task, bilingual participants filled in a language history questionnaire (in their L1) which included questions about, for example, the age of L2 acquisition, language use, other languages known to the subject and other variables that can influence language proficiency and experimental results. The results of this questionnaire can be seen in the participants section (cf. section 4.2.1, Table 4.1 and Table 4.3).

Procedure PNT

As shown in Figure 4.1, if the PNT was their first session, bilingual participants filled in the DIALANG test to assess their language competence prior to the experiment. This was done in order to assure that subjects had the required level to participate.

Next, participants read the instructions in their L2 (in order to increase the level of activation of the L2), and questions were answered by the instructor. A short practice phase followed so that participants could get used to the procedure. In preparation for the practice phase, participants received a catalogue of the pictures used in the practice phase and their corresponding names. They were told to study and remember the names and to only use those names in the practice phase. Prior to the experiment, participants received a catalogue of the pictures used in the experiment, with the same instructions.

The two naming or phrase type conditions (bare noun vs. NP) were administered in pseudorandomized order to each participant. Participants received a cue prior to each picture as to whether they should name the picture with or without the definite determiner. When the participants were to name the picture without the determiner, a "0" appeared in the center of the screen. For the NP condition, the cue consisted of two low dashes ("_ _") for the Spanish version because Spanish determiners have two letters (*el* or *la*). For the German version, the cue consisted of three low dashes ("_ _ _") because German determiners have three letters (*der*, *die*, *das*). Participants were instructed to name the pictures that would appear on the screen as quickly and correctly as possible and to avoid coughs, false starts, and hesitations (e.g., "uhmm"). The experiment was divided into two blocks between which participants could take a break as long as they wished. Each block started with two warming-up items which were not analyzed. It took about 15 - 20 minutes to complete the PNT.

The procedure was the same as the procedure employed by Lemhöfer, Spalek, and Schriefers (2008), with one exception. In Lemhöfer et al. (2008), production of NPs and bare nouns was blocked, here it was randomized in order to avoid the usage of strategies and to prevent subjects from settling into a response rhythm (cf. Bates et al., 2003; Szekely et al., 2005). At the beginning of each trial the naming cue was presented for 800 ms instead of a fixation cross. After a blank screen of 200 ms, the picture was presented for 2500 ms with a response timeout of 3000 ms after onset. The inter-trial interval was 750 ms.

RTs were recorded automatically by DMDX but response onsets were edited manually prior to data analysis with DMDX Check Vocal software (Protopapas, 2007). This was necessary as DMDX does not always record the voice onset times correctly, due to noises other than voice onsets, such as coughs, loud breathing, or swallowing. Moreover, because Check Vocal does not offer any error coding options, the different error types were recorded in an Excel table. This was done in order to allow for a more fine-grained analysis of error rates later on. Coded error types were phrase type errors (i.e., naming the picture without a determiner, while it was requested or vice versa), determiner errors, wrong picture names, failures to respond, responses that were too slow to be completely recorded, self-corrections, and hesitations. These errors were excluded from the RT analysis.

4.2.4 Data analysis

RTs to the PNT and LDT were analyzed for all experimental and control groups. Analyses of variance (GLM with repeated measures) were conducted on both participant and item mean RTs. In the analyses of the bilingual groups, the factors included were Gender Compatibility (congruent, incongruent, and incongruent neutral condition) and Phrase Type (PNT; bare noun naming vs. NP naming) or Prime Type (LDT; no determiner prime vs. definite determiner prime), Level (according to DIALANG test; low (A1-B1) vs. high (B2-C2)) and Task Order (LDT first–PNT second vs. PNT first–LDT second). In the subject analysis, the factors Gender Compatibility and Phrase Type/ Prime Type were within-subjects factors, while Level and Task Order were between-subjects factors. In the item analysis, the factors Phrase Type/Prime Type, Level, and Task Order were within-items factors and Gender Compatibility was a between-items factor. Follow-up analyses were conducted if an effect or interaction of interest was significant at least by participants or by items[22]. Greenhouse-Geisser corrections (Greenhouse-Geisser, 1959) were applied in cases where sphericity could not be assumed. Reported degrees of freedom are reported uncorrected but *p*-values are reported corrected.

For the monolingual groups, only the factors Gender Compatibility and Phrase Type/Prime Type applied, because all the monolingual subjects participated in only one of the tasks, either in the PNT or the LDT (cf. Participants section 4.2.1). In the analysis of the bare noun effect with the monolingual control group and bilingual group together, the bilingual group was split up according to proficiency levels in order to be able to investigate differences regarding proficiency. Hence, two separate analyses were conducted, one comparing the bilingual low-proficient group to the control group and another one comparing the bilingual high-proficient group to the control group. Here, the factors Experiment Group and Gender Compatibility (only bare nouns) were included.

Cognate Status was not included as a factor in the analyses because of the unequal amount of noncognates and cognates in the experiment (cf. section 4.2.2). Instead, three separate analyses were always conducted: one overall analysis with data pooled for noncognates and cognates, and two individual analyses for noncognates and cognates each.

Error rates were only analyzed for the PNT and offline gender assignment task because overall error rates in the LDT were very low (cf. section 4.8). The analysis of error rates was carried out in a similar fashion to the analyses of RTs, but only Gender Compatibility and Level were included as factors.

Items with errors in the translation task (different translations, wrong translations, and items of which the translation was unknown; cf. section 4.2.3) were excluded from the analyses on an individual basis as no information on gender transfer processes could be derived from them. Erroneous responses to the LDT (wrong key presses and failures to respond) and PNT (cf. section Procedure PNT in section 4.2.3) were also removed from each analysis. Outliers, that is, RTs deviating more than two standard deviations (SDs) from the participant and item conditions means were excluded as well. The number of errors and outliers is given before each analysis.

When reporting the results of the PNT and LDT, I will first look at the results of the monolingual control groups in order to check for possible artifacts caused by the experimental design[23]. If the interaction of Gender Compatibility with Phrase Type was not significant in the monolingual analysis, the data of the bilingual experimental group were analyzed regarding the NP gender interference effect. In the end, I will discuss the analysis of the bare noun effect in the bilingual group (split up by proficiency level) together with the monolingual control group. If, on the other hand, the relevant interaction of Gender Compatibility with Phrase Type was significant in the control group analysis, the data of the bilingual group were not analyzed because in this case, an interaction effect of Gender Compatibility with Phrase Type would not be informative. Instead, the bilingual data were analyzed in comparison with the monolingual control group data also regarding the NP effect. This was only the case for the PNT in German.

[22] Naturally, if an effect is only significant in one of the analyses, it has to be interpreted with caution.
[23] If item-matching was successful will be tested in the monolingual control groups of the LDT.

In the following section, first the analysis and results of the PNT, then the results of the offline gender assignment task and finally the analysis and results of the LDT will be reported. Across all tasks, first the analysis of the Spanish version of the task (Spanish monolingual control group and German bilingual participants) and then the analysis of the German version of the task (German monolingual control group and Spanish bilingual participants) will be described.

4.3 Results PNT

4.3.1 Results PNT in Spanish

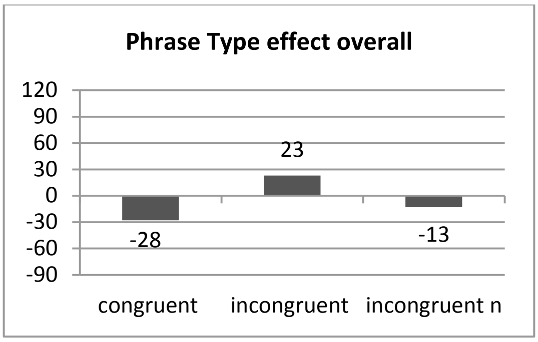
Results Spanish monolingual control group

The analysis was based on 20 subjects. There were 3.7 % (absolute count 85) phrase type errors (i.e., naming the picture without determiner, while the determiner was requested or vice versa), no wrong determiners, 3.9 % (88) wrong picture names, 1.5 % (35) failures to respond, and < 0.1 % (1) hesitations. These erroneous responses to the PNT were removed from the analysis. There were no outliers.

An overview of the RTs across Gender Compatibility and Phrase Type conditions and of the Phrase Type effect is given in Table 4.9 and Figure 4.2.

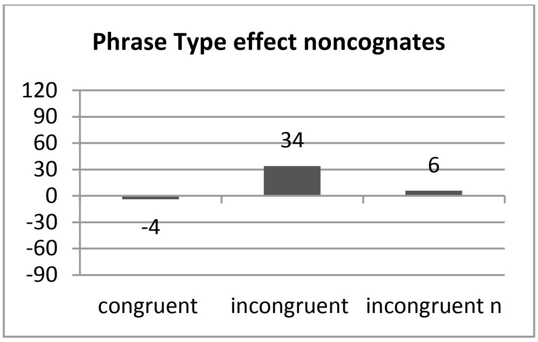
Overall

	Bare noun	NP
Congruent	981 (325)	953 (303)
Incongruent	984 (330)	1,007 (328)
Incongruent n	1,021 (334)	1,008 (333)



Noncognates

	Bare noun	NP
Congruent	957 (303)	953 (299)
Incongruent	971 (308)	1,005 (337)
Incongruent n	984 (315)	990 (350)



Cognates

	Bare noun	NP
Congruent	1,022 (357)	954 (313)
Incongruent	1,008 (367)	1,010 (313)
Incongruent n	1,082 (356)	1,040 (300)

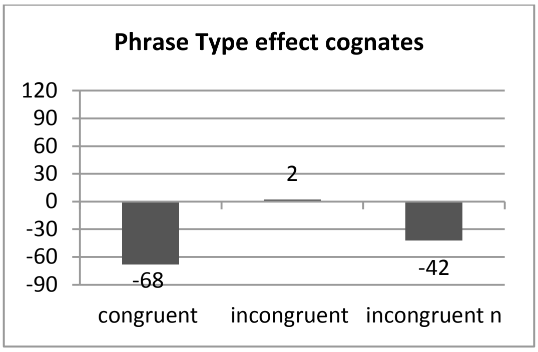


Table 4.9 RTs and standard deviations (in parentheses) for pooled cognates and noncognates (overall), for noncognates and cognates across Gender Compatibility and Phrase Type conditions.

Figure 4.2 The obtained Phrase Type effect for pooled cognates and noncognates (overall), for noncognates and cognates across Gender Compatibility conditions (Phrase Type effect = mean RTs of NP condition minus mean RTs of bare noun condition).

Statistical analyses

The results of the three analyses can be seen in Table 4.10. There were a few significant effects: There was a significant effect of Gender Compatibility in the overall and cognate F_1-analysis and this effect was also marginally significant in the F_1-analysis of noncognates.

		F1			***F2***		
		df	***F***	***p***	***df***	***F***	***p***
Overall	**Gender Compatibility**	2, 38	4.459	.028*	2, 111	.832	.438
	Phrase Type	1, 19	.124	.728	1,111	.438	.509
	Gender Compatibility * Phrase Type	2, 38	.982	.384	2, 111	1.140	.323
Noncognates	**Gender Compatibility**	2, 38	2.907	.067	2, 69	.489	.615
	Phrase Type	1, 19	.246	.626	1, 69	.376	.542
	Gender Compatibility * Phrase Type	2, 38	.363	.698	2, 69	.752	.475
Cognates	**Gender Compatibility**	2, 38	4.512	.017*	2, 39	.446	.643
	Phrase Type	1, 19	3.605	.073	1, 39	2.300	.137
	Gender Compatibility * Phrase Type	2, 38	.931	.403	2, 39	.433	.652

Table 4.10 Results of the F_1- and F_2-analyses, overall, for noncognates and cognates, with the factors Gender Compatibility and Phrase Type. Significant results are marked with an asterisk.

Summary

The analysis revealed two significant results: There was a significant effect of Gender Compatibility in the overall and cognate F_1-analysis. When looking at RTs across conditions it became apparent that RTs for the congruent and incongruent neutral condition were faster when pictures were named with a determiner than without, both overall and for cognates. In all three cases (overall, for noncognates and for cognates), RTs were slowest in the incongruent condition. This is unexpected as for the control group there should have been no differences between the conditions. However, since the relevant interaction of Gender Compatibility with Phrase Type was not significant in any of the analyses, the analyses of the bilingual group were conducted as planned. The results of the analysis of the bilingual group are summarized next.

Results bilingual German group

The analysis was based on 40 subjects. There were 19 high-proficient subjects and 21 low-proficient subjects. 21 of them had participated first in the LDT and 19 first in the PNT. Translation errors and unknown items as assessed in the translation task were removed per participant before analysis. In the PNT, there were 5.8 % (absolute count 264) phrase type errors (i.e., naming the picture without a determiner, while it was requested or vice versa), 1.9 % (85) wrong determiners, 5.3 % (245) wrong picture names, 4.6 % (209) failures to respond and no hesitations. These erroneous responses to the PNT were removed from the analysis. No outliers were observed. Furthermore, five items (*nudo–Knoten, florero–Vase, fusil–Gewehr, sal–Salz, orquesta–Orchester*) were excluded from the analysis due to a lack of observations in some conditions for the item analysis. 109 items were left in the analysis, 69 (out of 72) noncognates and 40 (out of 42) cognates. In each of the Gender Compatibility conditions, the following amount of items was left: congruent condition = 37 items, incongruent condition = 37 items, incongruent neutral condition = 35 items. For noncognates, there were 23 items left in each condition. For cognates, there were 14 items in the congruent condition, 14 in the incongruent condition and 12 in the incongruent neutral condition.

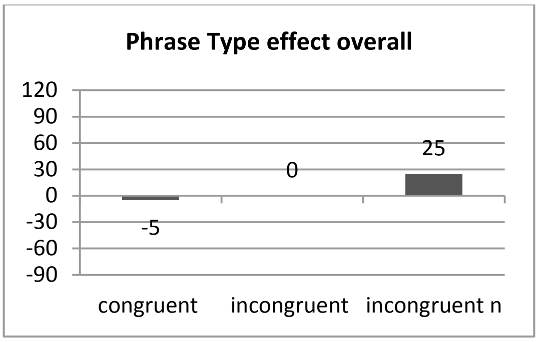
An overview of the RTs is given in Table 4.11.

			PNT 1st			PNT 2nd		
			Bare noun	**NP**	**Phrase Type effect**	**Bare noun**	**NP**	**Phrase Type effect**
Low-proficient group	**Noncognates**	**Congruent**	1,231 (382)	1,232 (348)	+ 1	1,143 (361)	1,174 (374)	+ 31
		Incongruent	1,186 (330)	1,180 (340)	- 6	1,192 (390)	1,181 (360)	- 11
		Incongruent n	1,227 (437)	1,257 (349)	+ 30	1,187 (335)	1,197 (353)	+ 10
	Cognates	**Congruent**	1,177 (374)	1,145 (306)	- 32	1,128 (395)	1,206 (427)	+ 78
		Incongruent	1,121 (314)	1,223 (338)	+ 2	1,117 (396)	1,185 (347)	+ 68
		Incongruent n	1,188 (304)	1,188 (317)	0	1,251 (368)	1,247 (359)	- 4
			PNT 1st			**PNT 2nd**		
			Bare noun	**NP**	**Phrase Type effect**	**Bare noun**	**NP**	**Phrase Type effect**
High-proficient group	**Noncognates**	**Congruent**	1,094 (356)	1,018 (242)	- 76	1,204 (344)	1,167 (304)	- 37
		Incongruent	1,108 (364)	1,040 (255)	- 68	1,198 (364)	1,114 (278)	- 84
		Incongruent n	1,095 (293)	1,102 (353)	+ 7	1,161 (334)	1,227 (312)	- 66
	Cognates	**Congruent**	1,025 (364)	1,035 (318)	+ 10	1,155 (353)	1,140 (291)	- 15
		Incongruent	1,062 (311)	1,150 (358)	+ 88	1,187 (334)	1,194 (323)	+ 7
		Incongruent n	999 (235)	1,121 (349)	+ 122	1,249 (355)	1,191 (252)	- 58

Table 4.11 Overview of the RTs and standard deviations (in parentheses) per Level (low- vs. high-proficient group), Cognate Status (noncognates vs. cognates), Gender Compatibility condition (congruent, incongruent, and incongruent neutral condition), Task Order (PNT first vs. PNT second), and Phrase Type (bare noun vs. NP). Phrase Type effect = mean RTs of NP condition minus mean RTs of bare noun condition.

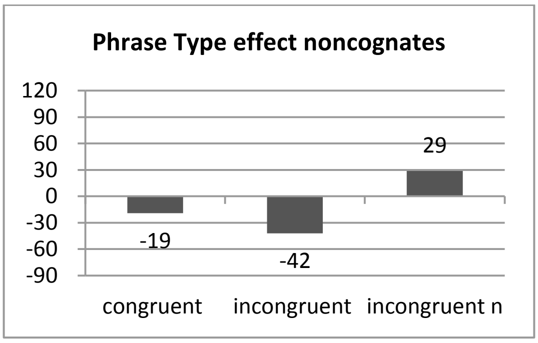
As described in section 4.1.2, I expected either a small facilitation or no effect for the primed congruent condition and significant inhibition for the primed incongruent conditions, compared to the unprimed conditions. As can be seen in Table 4.12, the observed effect looked different. There was no consistent Phrase Type effect across Cognate Status conditions.

		Bare noun	**NP**
Overall	**Congruent**	1,143 (368)	1,138 (335)
	Incongruent	1,148 (357)	1,148 (326)
	Incongruent n	1,165 (343)	1,190 (340)



Noncognates

	Bare noun	NP
Congruent	1,161 (362)	1,142 (328)
Incongruent	1,168 (365)	1,126 (315)
Incongruent n	1,163 (348)	1,192 (347)



Cognates

	Bare noun	NP
Congruent	1,117 (375)	1,131 (347)
Incongruent	1,115 (342)	1,185 (341)
Incongruent n	1,167 (334)	1,186 (327)

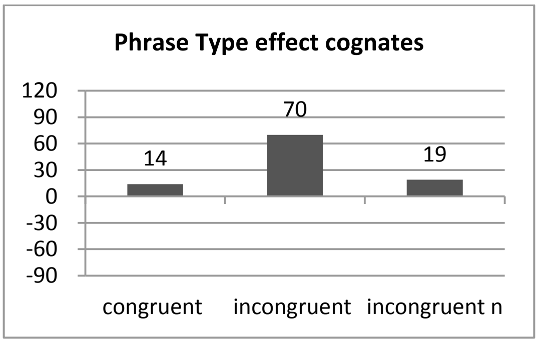


Table 4.12 RTs and standard deviations (in parentheses) for pooled cognates and noncognates (overall), for noncognates and cognates across Gender Compatibility and Phrase Type conditions.

Figure 4.3 The obtained Phrase Type effect for pooled cognates and noncognates (overall), for noncognates and cognates across Gender Compatibility conditions (Phrase Type effect = mean RTs of NP condition minus mean RTs of bare noun condition).

As can be seen in Table 4.13, in the overall analysis the interaction of Gender Compatibility with Phrase Type was not significant and there was no significant higher-order interaction including this interaction. Neither was there a significant interaction when the analysis was split up according to Cognate Status. In the following, the data and effects are described in more detail.

Overall analysis: As shown in Table 4.13, RTs across Gender Compatibility conditions were largely similar, except for the incongruent neutral condition which had slightly longer naming times (congruent: 1,141 ms, SD 352 ms; incongruent: 1,148 ms, SD 342 ms; incongruent neutral: 1,177 ms, SD 342 ms). The effect of Gender Compatibility was significant in the F_1-analysis but not in the F_2-analysis (Table 4.13). On the whole, the need to produce the picture name with the correct determiner had little effect on RTs. Naming with determiners (1,158 ms, SD 334 ms) slowed down RTs by only 6 ms on average compared to bare noun naming (1,153 ms, SD 356 ms), which was not significant. High-proficient subjects had faster RTs (1,121 ms, SD 326 ms) than low-proficient subjects (1,190 ms, SD 362 ms), but this difference was only significant in the F_2-analysis[24]. Regarding task order effects, subjects performed on average 52 ms more slowly if the PNT was the second task (1,181 ms, SD 349 ms), compared to 1,129 ms (SD 340 ms), when it was the first task. This difference was only significant in the F_2-analysis. The interaction of Level with Task Order was also significant in the F_2-analysis.

[24] Note that nevertheless, as mentioned before, the difference between the two Levels (high- and low-proficient) was significant regarding the CEFR levels (χ^2 =44.000, df = 4, $p < 0.001$).

		F1			F2		
		df	***F***	***p***	***df***	***F***	***p***
Overall	**Gender Compatibility**	2, 72	5.118	.008*	2, 106	1.157	.319
	Phrase Type	1, 36	.001	.978	1, 106	.014	.905
	Gender Compatibility * Phrase Type	2, 72	.782	.461	2, 106	.332	.719
	Level	1, 36	1.979	.168	1, 106	45.350	< .001*
	Task Order	1, 36	1.277	.266	1, 106	17.698	< .001*
	Level * Task Order	1, 36	2.229	.144	1, 106	29.039	< .001*
Noncognates	**Gender Compatibility**	2, 72	4.970	.010*	2, 66	1.257	.291
	Gender Compatibility * Level * Task Order	2, 72	3.208	.046*	2, 66	1.464	.239
	Phrase Type	1, 36	0.000	.324	1, 66	.966	.329
	Gender Compatibility * Phrase Type	2, 72	1.078	.346	2, 66	1.463	.239
	Level	1, 36	2.393	.131	1, 66	26.250	< .001*
	Task Order	1, 36	.574	.454	1, 66	4.876	.031*
	Level * Task Order	1, 36	2.685	.110	1, 66	21.867	< .001*

Cognates	**Gender Compatibility**	2, 72	2.416	.096	2, 37	.156	.856
	Gender Compatibility * Level	2, 72	1.061	.351	2, 37	2.564	.091
	Phrase Type	1, 36	2.161	.150	1, 37	2.749	.106
	Phrase Type * Task Order	1, 36	3.029	.090	1, 37	1.004	.323
	Phrase Type * Level* Task Order	1, 36	3.154	.084	1, 37	1.517	.226
	Gender Compatibility * Phrase Type	2, 72	.505	.570	2, 37	.664	.521
	Level	1, 36	1.127	.295	1, 37	22.205	< .001*
	Task Order	1, 36	2.247	.143	1, 37	17.733	< .001*
	Level * Task Order	1, 36	1.422	.241	1, 37	7.009	.012*

Table 4.13 Results of the overall, noncognate, and cognate F_1- and F_2-analysis with the factors Gender Compatibility, Phrase Type, Level, and Task Order. Effects are only displayed if they are (a) theoretically important (i.e., main effects and interaction effect of Gender Compatibility and Phrase Type), (b) if their *p*-value is < .10. Main effects and interactions that are not relevant with regard to the predictions and with a *p*-value > .10 are not displayed.

Noncognate analysis (cf. Table 4.13): As in the overall analysis, there was a significant main effect of Level in the F_2-analysis, a significant main effect of Task Order in the F_2-analysis and an interaction effect of Level with Task Order significant in the F_2-analysis. There was also a significant effect of Gender Compatibility for F_1 and a significant interaction of Gender Compatibility with Level and Task Order for F_1.

Cognate Analysis (cf. Table 4.13): As in the two previous analyses, there was a significant main effect of Level and of Task Order both in the F_2-analysis as well as an interaction effect of Level with Task Order which was also only significant in the F_2-analysis.

Summary

There was no interaction effect of Gender Compatibility with Phrase Type in any of the three analyses. Furthermore, the RT patterns across Gender Compatibility conditions were different from the expected pattern and inconsistent between Cognate Status conditions. There was a significant effect of Level in the F_2-analyses across all three analyses as high-proficient subjects had faster RTs than low-proficient subjects.

Analysis of bilingual German group with monolingual Spanish control group

This analysis was conducted in order to investigate the bare noun gender interference effect. It is based on the same participants and items as the previous analysis.

Bare noun effect

The descriptive data of bare noun naming times and the bare noun effect across condition shown in Table 4.14 and Figure 4.4, indicate that there does not seem to be a consistent Gender Compatibility effect across the three analyses, neither for the low-proficient nor for the high-proficient group. None of the RT patterns corresponds to the expected bare noun effect, that is, faster RTs for the congruent condition than for the two incongruent conditions.

Overall

	Mono-linguals	Bilinguals	
		Low	High
Congruent	981 (325)	1,167 (377)	1,121 (358)
Incongruent	984 (330)	1,162 (363)	1,134 (351)
Incongruent n	1,021 (334)	1,210 (365)	1,121 (315)

Noncognates

	Mono-linguals	Bilinguals	
		Low	High
Congruent	957 (303)	1,181 (371)	1,143 (354)
Incongruent	971 (308)	1,189 (364)	1,148 (366)
Incongruent n	984 (315)	1,203 (379)	1,125 (313)

Cognates

	Mono-linguals	Bilinguals	
		Low	High
Congruent	1,022 (357)	1,148 (386)	1,085 (363)
Incongruent	1,008 (367)	1,119 (359)	1,111 (325)
Incongruent n	1,082 (356)	1,223 (340)	1,114 (320)

Table 4.14 RTs and standard deviations (in parentheses) of the monolingual group and low- and high-proficient bilingual groups for pooled cognates and noncognates (overall), for noncognates and cognates for bare noun naming across Gender Compatibility conditions.

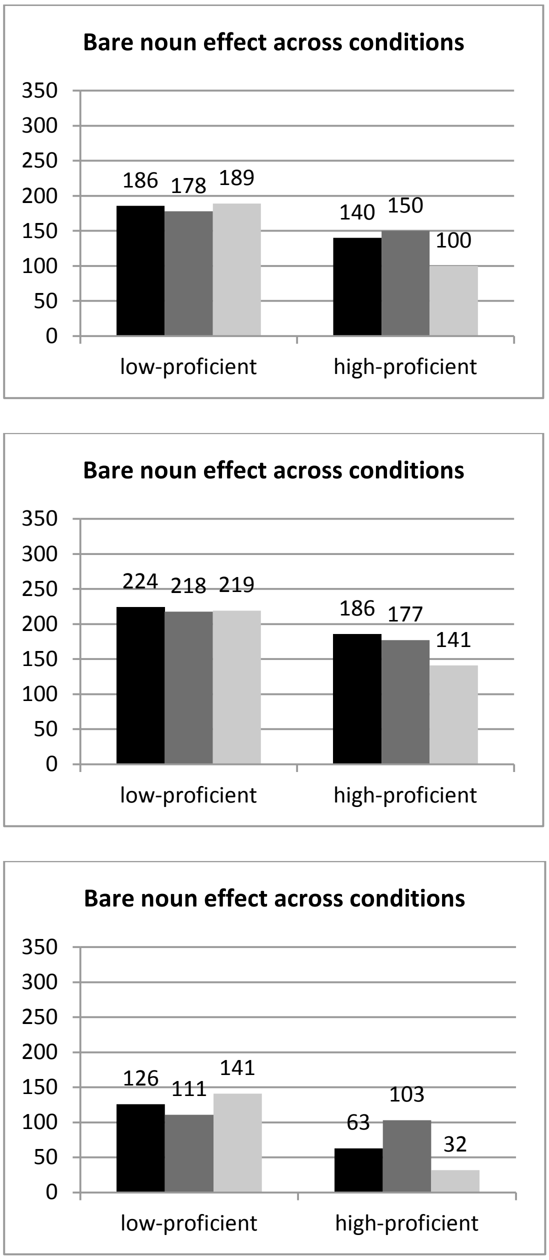


Figure 4.4 The obtained bare noun effect for pooled cognates and noncognates (overall), for noncognates and cognates across Gender Compatibility conditions in the two proficiency groups of the bilingual group. (Bare noun effect = mean RTs of bilingual group minus mean RTs of monolingual group).

For the low-proficient bilingual group, statistical analyses revealed that the interaction of Gender Compatibility (only bare nouns) with Experiment Group was not significant, neither in the overall analysis ($F_1(2, 78) = 0.219$, $p = 0.804$; $F_2(2, 106) = 0.142$, $p = 0.868$), nor in the noncognate analysis ($F_1(2, 78) = 0.249$, $p = 0.746$; $F_2(2, 66) = 0.021$, $p = 0.979$), nor cognate analysis ($F_1(2, 78) = 0.024$, $p = 0.976$; $F_2(2, 37) = 0.229$, $p = 0.796$).

For the high-proficient bilingual group, statistical analyses revealed that the interaction of Gender Compatibility (only bare nouns) with Experiment Group was not significant, neither in the overall analysis ($F_1(2, 74) = 0.629$, $p = 0.536$; $F_2(2, 106) = 0.448$, $p = 0.640$), nor in the noncognate analysis ($F_1(2, 74) = 0.985$, $p = 0.378$; $F_2(2, 66) = 0.375$, $p = 0.689$), nor cognate analysis ($F_1(2, 74) = 1.105$, $p = 0.337$; $F_2(2, 37) = 0.425$, $p = 0.657$).

Summary

The analysis comparing the bilingual experimental with the monolingual control group failed to reveal a significant difference between Gender Compatibility conditions for bare nouns. Hence, there is no evidence for a bare noun gender interference effect.

4.3.2 Results PNT in German

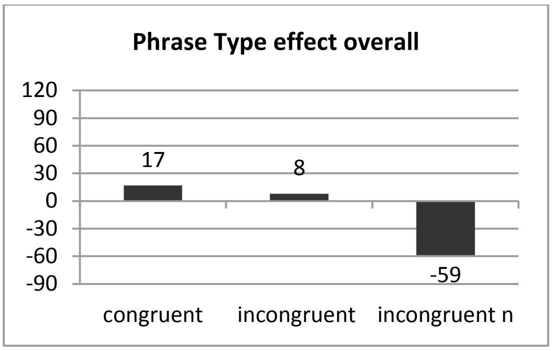
Results German monolingual control group

The analysis was based on 20 subjects. There were 2.8 % (absolute count 64) phrase type errors (i.e., naming the picture without determiner, while the determiner was requested or vice versa), no wrong determiners, 3.5 % (79) wrong picture names, 0.6 % (14) failures to respond and 0.4 % (9) hesitations. These erroneous responses to the PNT were removed from the analysis. There were no outliers.

An overview of the RTs across Gender Compatibility and Phrase Type conditions and of the Phrase Type effect is given in Table 4.15 and Figure 4.5.

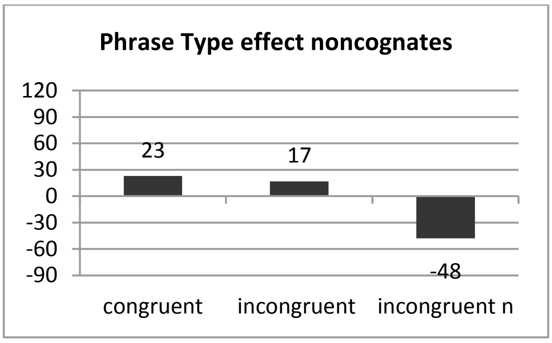
Overall

	Bare noun	NP
Congruent	929 (309)	946 (305)
Incongruent	972 (311)	980 (337)
Incongruent n	1,001 (289)	942 (269)



Noncognates

	Bare noun	NP
Congruent	911 (303)	934 (292)
Incongruent	950 (290)	967 (349)
Incongruent n	968 (271)	920 (256)



Cognates	Bare noun	NP
Congruent	958 (317)	968 (327)
Incongruent	1,011 (344)	1,005 (314)
Incongruent n	1,063 (310)	981 (288)

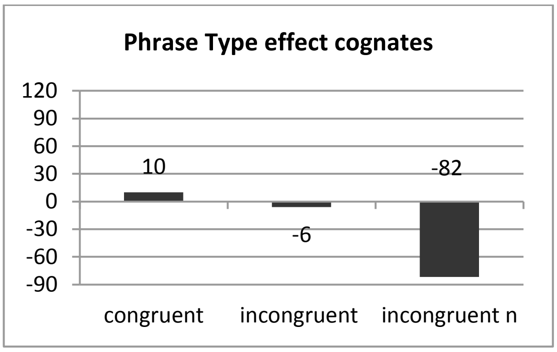


Table 4.15 RTs and standard deviations (in parentheses) for pooled cognates and noncognates (overall), for noncognates and cognates across Gender Compatibility and Phrase Type conditions.

Figure 4.5 The obtained Phrase Type effect for pooled cognates and noncognates (overall), for noncognates and cognates across Gender Compatibility conditions (Phrase Type effect = mean RTs of NP condition minus mean RTs of bare noun condition).

Statistical analyses

The results of the three analyses can be seen in Table 4.16. There were a few significant effects: There was a significant effect of Gender Compatibility in the overall and cognate F_1-analysis and the interaction of Gender Com**patibility with Phrase Type was significant for F_1 and F_2 in the overall analysis as well as for F_1 in the** noncognate analysis.

		F1			*F2*		
		df	*F*	*p*	*df*	*F*	*p*
Overall	**Gender Compatibility**	2, 38	4.709	.035*	2, 111	.093	.409
	Phrase Type	1, 19	.131	.721	1, 111	.550	.460
	Gender Compatibility * Phrase Type	2, 38	3.581	.038*	2, 111	4.409	.014*
oncognates	**Gender Compatibility**	2, 38	2.597	.113	2, 69	.397	.674
	Phrase Type	1, 19	.015	.905	1, 69	.000	.991
	Gender Compatibility * Phrase Type	2, 38	3.347	.046*	2, 69	2.177	.121
Cognates	**Gender Compatibility**	2, 38	3.649	.035*	2, 39	.590	.559
	Phrase Type	1, 19	1.128	.301	1, 39	1.372	.249
	Gender Compatibility * Phrase Type	2, 38	1.283	.289	2, 39	2.277	.116

Table 4.16 Results of the F_1- and F_2-analyses, overall, for noncognates and cognates, with the factors Gender Compatibility and Phrase Type. Significant results are marked with an asterisk.

Summary

The analysis revealed a few significant effects: There was a significant effect of Gender Compatibility in the overall and cognate F_1-analysis. Furthermore, the interaction of Gender Compatibility with

Phrase Type was significant in the F_1- and F_2-analysis in the overall analysis and in the noncognate F_1-analysis. When looking at RTs across conditions, it becomes apparent that the incongruent neutral condition exhibits a strong facilitation when named with a determiner and behaves very differently from the other conditions in all three analyses. Because of this significant interaction of Gender Compatibility with Phrase Type in the control group, the analysis of the NP effect in the bilingual group was conducted in comparison with the monolingual control group, just as the analysis of the bare noun effect.

Analysis of bilingual Spanish group with monolingual German control group

The analysis was based on 30 bilingual and 20 monolingual subjects. Of the bilingual subjects, there were 18 high-proficient subjects and 12 low-proficient subjects. 13 of them had participated first in the LDT and 17 first in the PNT. Translation errors and unknown items as assessed in the translation task were removed per participant before analysis. In the PNT, there were 4.4 % (absolute count 150) phrase type errors (i.e., naming the picture without a determiner while it was requested or vice versa), 9.3 % (318) wrong determiners, 4.2 % (142) wrong picture names, 0.5 % (18) responses that were too slow to be completely recorded, 3.7 % (125) failures to respond, 0.2 % (8) self-corrections and 0.2 % (6) hesitations. These erroneous responses to the PNT were removed from the analysis. No outliers were observed.

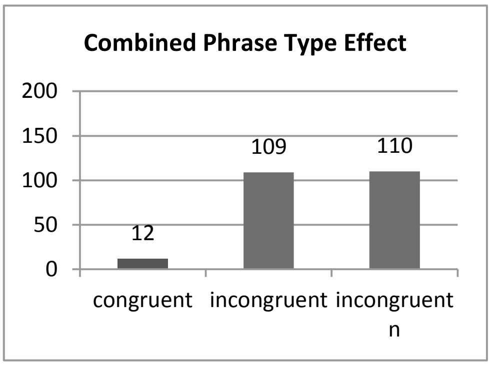
In the analysis of the NP effect with the low-proficient group, five items had to be removed (*rama–Ast*, *disco–CD*, *cuerda–Seil*, *verdura–Gemüse*, *aguacate–Avocado*) due to a lack of observations in some conditions for the item analysis. For better comparability between the two proficiency groups, these five items were also removed in the high-proficient analysis and also in the bare noun analyses. 109 items were left in each analysis, 68 (out of 72) noncognates and 41 (out of 42) cognates. In the congruent condition, 38 items were left and in the incongruent condition 35 items and in the incongruent neutral condition 36 items. For noncognates, there were 24 items in the congruent condition, 22 items in the incongruent condition and 22 items in the incongruent neutral condition. For cognates, there were 14 items left in the congruent and incongruent neutral condition and 13 in the incongruent condition.

NP effect – low-proficient group

Low-proficient bilingual group

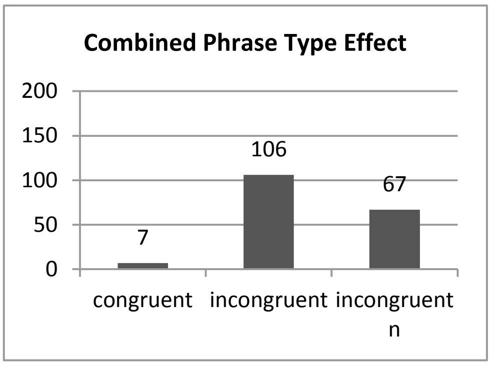
Overall

	Bare noun	NP	Phrase Type effect	Effect control group
Congruent	1,241 (390)	1,270 (349)	+ 29	+ 17
Incongruent	1,201 (350)	1,318 (382)	+ 117	+ 8
Incongruent n	1,253 (347)	1,304 (375)	+ 51	- 59



Noncognates

	bare noun	NP	Phrase Type effect	Effect control group
congruent	1,249 (396)	1,279 (343)	+ 30	+ 23
incongruent	1,225 (348)	1,348 (380)	+ 123	+ 17
Incongruent n	1,250 (358)	1,269 (360)	+ 19	- 48



Cognates

	Bare noun	NP	Phrase Type effect	Effect control group
Congruent	1,230 (382)	1,254 (362)	+ 24	+ 10
Incongruent	1,162 (352)	1,264 (383)	+ 102	- 6
Incongruent n	1,258 (332)	1,370 (396)	+ 112	- 82

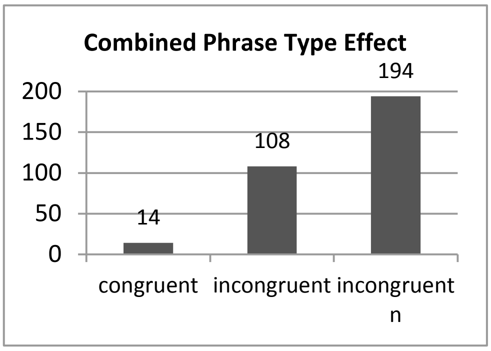


Table 4.17 RTs and standard deviations (in parentheses) of the low-proficient bilingual group for pooled cognates and noncognates (overall), for noncognates and cognates across Gender Compatibility and Phrase Type conditions. The bilingual and monolingual Phrase Type effects are also displayed.

Figure 4.6 The combined Phrase Type effect for pooled cognates and noncognates (overall), for noncognates and cognates across Gender Compatibility conditions (Combined Phrase Type effect = Phrase Type effect of bilingual group minus Phrase Type effect of monolingual group).

An overview of the RTs of the low-proficient bilinguals is given in Table 4.17. The “Combined Phrase Type Effect”, that is, the bilingual NP effect relative to the monolingual control group, is depicted in Figure 4.6. At face value, it seemed that there was an inhibition effect for the two incongruent conditions.

However, as can be seen in Table 4.18, in the overall and cognate analysis the interaction of Experiment Group with Gender Compatibility and Phrase Type was only marginally significant in the F_2-analysis. In the noncognate analysis, the interaction was not significant in either the F_1- or F_2-analysis. In the following, the data and effects are described in more detail.

Overall analysis: The results are shown in Table 4.18. There was a significant effect of Experiment Group as monolinguals had significantly faster naming times (962 ms, SD 305 ms) than (low-proficient) bilinguals (1,260 ms, SD 367 ms). RTs across Gender Compatibility conditions were largely similar (congruent: 1,038 ms, SD 361 ms; incongruent: 1,057 ms, SD 360 ms; incongruent neutral: 1,065 ms, SD 337 ms). The effect of Gender Compatibility was not significant. The effect of Phrase Type was significant in the F_2-analysis and marginally significant in the F_1-analysis. Naming times in the NP condition (1,048 ms, SD 357 ms) were slightly longer than in the bare noun condition (1,048 ms, SD 357 ms). The interaction of Phrase Type with Experiment Group was significant in the F_1- and

F_2-analysis. Follow-up analyses revealed that there was a significant effect of Phrase Type in the bilingual group ($F_1(1, 11) = 5.238$, $p = 0.043$; $F_2(1, 106) = 14.753$, $p < 0.001$) but not for the monolingual group ($F_1(1, 19) = 0.101$, $p = 0.754$; $F_2(1, 106) = 2.136$, $p = 0.147$). Descriptive data showed that the monolinguals had slightly faster naming times in the NP condition (956 ms, SD 305 ms) than the bare noun condition (967 ms, SD 304 ms). Low-proficient bilinguals, on the other hand, had much slower naming times in the NP condition (1,296 ms, SD 368 ms) than in the bare noun condition (1,233 ms, SD 364 ms).

			F1			***F2***		
			df	***F***	***p***	***df***	***F***	***p***
Low-proficient bilingual group and monolingual control group	**Overall**	**Gender Compatibility**	2, 60	2.024	.141	2, 106	.449	.639
		Phrase Type	1, 30	3.580	.068	1, 106	9.574	.003*
		Experiment Group * Phrase Type	1, 30	5.334	.028*	1, 106	16.063	< .001*
		Experiment Group * Gender Compatibility * Phrase Type	2, 60	1.100	.339	1, 106	2.758	.068
		Experiment Group	1, 30	36.458	< .001*	1, 106	491.835	< .001*
	Noncognates	**Gender Compatibility**	2, 60	.648	.496	2, 65	.001	.999
		Phrase Type	1, 30	3.324	.078	1, 65	5.148	.027*
		Experiment Group * Phrase Type	1, 30	2.857	.101	1, 65	6.219	.015*
		Experiment Group * Gender Compatibility * Phrase Type	2, 60	.685	.485	2, 65	.664	.518
		Experiment Group	1, 30	42.347	< .001*	1, 65	380.377	< .001*
	Cognates	**Gender Compatibility**	2, 60	4.062	.022*	2, 38	.886	.420
		Phrase Type	1, 30	3.187	.084	1, 38	4.225	.047*
		Experiment Group * Phrase Type	1, 30	9.233	.005*	1, 38	11.791	.001*
		Experiment Group * Gender Compatibility * Phrase Type	2, 60	.479	.622	2, 38	3.124	.055
		Experiment Group	1, 30	24.617	< .001*	1, 38	132.768	< .001*

Table 4.18 Results of the overall F_1- and F_2-analysis for the low-proficient bilingual and monolingual control group with the factors Gender Compatibility, Phrase Type, and Experiment Group. Effects are only displayed if they are (a) theoretically important (i.e., all main effects and the interaction effect of Gender Compatibility with Phrase Type and Experiment Group), (b) if their *p*-value is < .10. Main effects and interactions that are not relevant with regard to the predictions and with a *p*-value > .10 are not displayed.

Noncognate analysis (cf. Table 4.18): The effect of Phrase Type was significant in the F_2-analysis and marginally significant in the F_1-analysis. The interaction of Experiment Group with Phrase Type was significant in the F_2-analysis. As in the overall analysis, there was a significant main effect of Experiment Group.

Cognate Analysis (cf. Table 4.18): There was a significant effect of Gender Compatibility in the F_1-analysis. The effect of Phrase Type was significant for F_2 and marginally significant in the F_1-analysis.

There was a significant interaction of Experiment Group with Phrase Type, as in the overall analysis. The effect of Experiment Group was also significant.

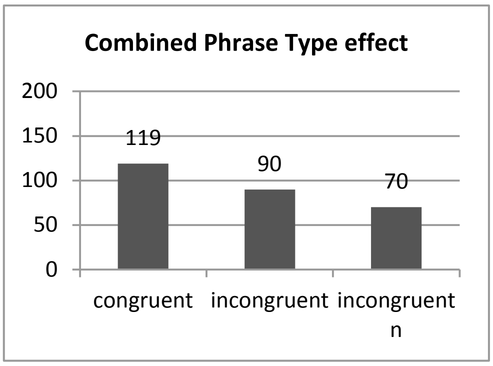
NP-effect high-proficient group

An overview of the RTs of the high-proficient bilinguals is given in Table 4.19. The "Combined Phrase Type Effect" is depicted in Figure 4.7. No consistent RT pattern across the three different analyses can be observed.

High-proficient bilingual group

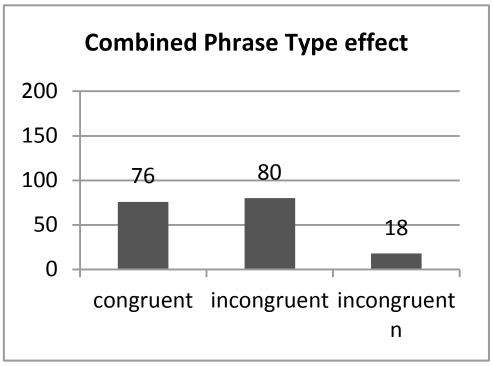
Overall

	Bare noun	NP	Phrase Type effect	Effect control group
Congruent	1,168 (353)	1,259 (400)	+ 91	- 28
Incongruent	1,194 (357)	1,307 (402)	+ 113	+ 23
Incongruent n	1,203 (354)	1,260 (407)	+ 57	- 13



Noncognates

	Bare noun	NP	Phrase Type effect	Effect control group
Congruent	1,182 (366)	1,254 (381)	+ 72	- 4
Incongruent	1,204 (367)	1,318 (394)	+ 114	+ 34
Incongruent n	1,190 (366)	1,214 (376)	+ 24	+ 6



Cognates

	Bare noun	NP	Phrase Type effect	Effect control group
Congruent	1,145 (331)	1,268 (434)	+ 123	- 68
Incongruent	1,178 (340)	1,289 (419)	+ 111	+ 2
Incongruent n	1,223 (332)	1,345 (448)	+ 122	- 42

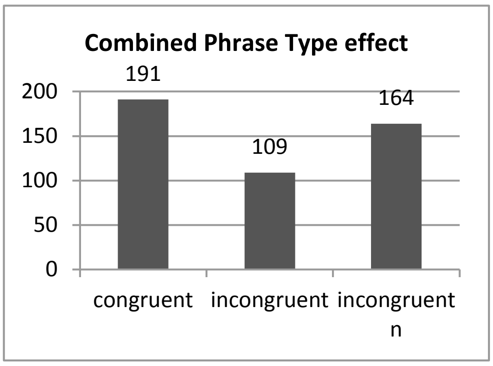


Table 4.19 RTs and standard deviations (in parentheses) of the high-proficient bilingual group for pooled cognates and noncognates (overall), for noncognates and cognates across Gender Compatibility and Phrase Type conditions. The bilingual and monolingual Phrase Type effects are also displayed.

Figure 4.7 The combined Phrase Type effect for pooled cognates and noncognates (overall), for noncognates and cognates across Gender Compatibility conditions (Combined Phrase Type effect = Phrase Type effect of bilingual group minus Phrase Type effect of monolingual group).

As can be seen in Table 4.20, in interaction of Experiment Group with Gender Compatibility and Phrase Type was not significant in any of the analyses. In the following, the data and effects are described in more detail.

Overall analysis: The results are shown in Table 4.20. There effect of Gender Compatibility was significant in the F_1-analysis. RTs were somewhat faster in the congruent (1,057 ms, SD 365 ms) than in the two incongruent conditions (incongruent: 1,090 ms, SD 374 ms; incongruent neutral: 1,081 ms, SD 350 ms). The effect of Phrase Type was significant in the F_1- and F_2-analysis. Overall RTs were somewhat slower in the NP naming condition than in the bare noun naming conditions (1,086 ms, SD 382 ms and 1,066 ms, SD 346 ms, respectively). There was also a significant interaction of Experiment Group with Phrase Type. Follow-up analyses revealed that there was a significant main effect of Phrase Type for the high-proficient bilingual group ($F_1(1, 17) = 14.947$, $p = 0.001$; $F_2(1, 106) = 40.429$, $p < 0.001$) but not for the monolingual group ($F_1(1, 19) = 0.101$, $p = 0.754$; $F_2(1, 106) = 2.136$, $p = 0.147$). Descriptive data showed that RTs between Phrase Type conditions differed only slightly for monolinguals (bare noun naming: 967 ms, SD 304 ms; NP naming: 956 ms, SD 305 ms). For bilinguals, on the other hand, naming times were much longer for the NP condition (1,275 ms, SD 403 ms) than for the bare noun condition (1,188 ms, SD 354 ms). The interaction of Gender Compatibility and Phrase Type was significant in the F_2-analysis. There was a significant main effect of Experiment Group as naming times of monolinguals (962 ms, SD 305 ms) were much faster than of high-proficient bilinguals (1,228 ms, SD 380 ms).

			F1			*F2*		
			df	***F***	***p***	***df***	***F***	***p***
High-proficient bilingual group and monolingual control group	**Overall**	**Gender Compatibility**	2, 72	5.390	.007*	2, 106	.393	.676
		Phrase Type	1, 36	7.424	.010*	1, 106	20.333	< .001*
		Experiment Group * Phrase Type	1, 36	10.282	.003*	1, 106	38.376	< .001*
		Gender Compatibility * Phrase Type	2, 72	1.629	.203	2, 106	3.672	.029*
		Experiment Group * Gender Compatibility * Phrase Type	2, 72	.675	.512	2, 106	.506	.604
		Experiment Group	1, 36	26.858	< .001*	1, 106	424.107	< .001*
	Noncognates	**Gender Compatibility**	2, 72	4.893	.010*	2, 65	.355	.702
		Phrase Type	1, 36	4.574	.039*	1,65	10.216	.002*
		Experiment Group * Phrase Type	1, 36	3.926	.055	1,65	15.610	< .001*

		Experiment Group * Gender Compatibility * Phrase Type	2, 72	.640	.530	2, 65	.393	.677
		Experiment Group	1, 36	29.052	< .001*	1,65	286.442	< .001*
	Cognates	**Gender Compatibility**	2, 72	5.511	.006*	2, 38	.355	.703
		Phrase Type	1, 36	7.959	.008*	1, 38	10.168	.003*
		Experiment Group * Phrase Type	1, 36	17.139	< .001*	1, 38	24.907	< .001*
		Experiment Group * Gender Compatibility * Phrase Type	2, 72	.858	.428	2, 38	.451	.640
		Experiment Group	1, 36	21.388	< .001*	1, 38	136.172	< .001*

Table 4.20 Results of the overall F_1- and F_2-analysis for the high-proficient bilingual and monolingual control group with the factors Gender Compatibility, Phrase Type, and Experiment Group. Effects are only displayed if they are (a) theoretically important (i.e., all main effects and the interaction effect of Gender Compatibility with Phrase Type and Experiment Group), (b) if their p-value is < .10. Main effects and interactions that are not relevant with regard to the predictions and with a p-value > .10 are not displayed.

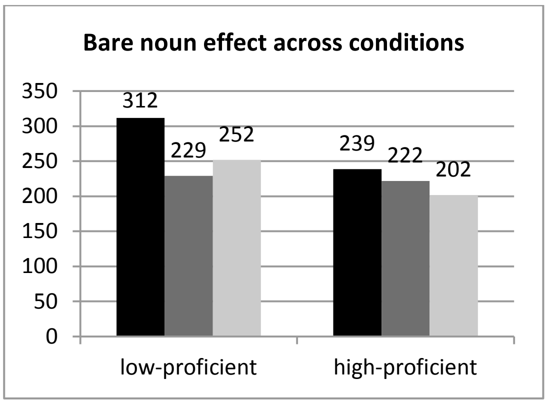
Noncognate analysis (cf. Table 4.20): The effect of Gender Compatibility was significant in the F_1-analysis but not significant in the F_2-analysis. The effect of Phrase Type was significant in the F_1-analysis and the F_2-analysis. The interaction of Experiment Group with Phrase Type was significant in the F_2-analysis and marginally significant in the F_1-analysis. As in the overall analysis, there was a significant main effect of Experiment Group.

Cognate Analysis (cf. Table 4.20): There was a significant effect of Gender Compatibility in the F_1-analysis. The effect of Phrase Type was significant in the F_1-analysis and the F_2-analysis. The interaction of Experiment Group with Phrase Type was also significant in the F_1-analysis and the F_2-analysis. As in the two other analyses, the main effect of Experiment Group was significant.

Bare noun effect

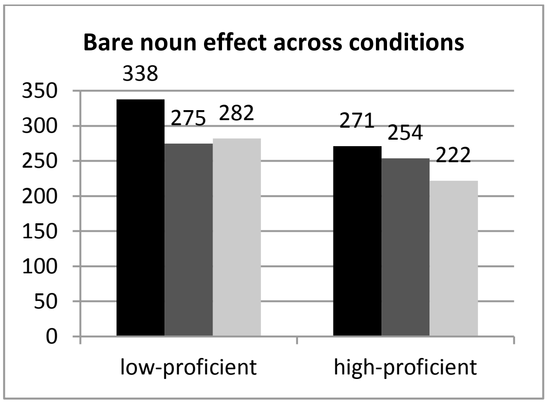
The descriptive data of bare noun naming times and the bare noun effect across condition shown in Table 4.21 and Figure 4.8 indicate that there does not seem to be a consistent Gender Compatibility effect across the three analyses, neither for the low-proficient nor for the high-proficient group. None of the RT patterns corresponds to the expected bare noun effect, that is, faster RTs for the congruent condition than for the incongruent conditions. For the low-proficient group, however, naming times are consistently longer for the congruent condition than for the incongruent condition.

		Mono-linguals	**Bilinguals**	
			Low	**High**
Overall	**Congruent**	929 (309)	1,241 (390)	1,168 (353)
	Incongruent	972 (311)	1,201 (350)	1,194 (357)
	Incongruent n	1,001 (289)	1,253 (347)	1,203 (354)



Noncognates

	Mono-linguals	Bilinguals	
		Low	High
Congruent	911 (303)	1,249 (396)	1,182 (366)
Incongruent	950 (290)	1,225 (348)	1,204 (367)
Incongruent n	968 (271)	1,250 (358)	1,190 (366)



Cognates

	Mono-linguals	Bilinguals	
		Low	High
Congruent	958 (317)	1,230 (382)	1,145 (331)
Incongruent	1,011 (344)	1,162 (352)	1,178 (340)
Incongruent n	1,063 (310)	1,258 (332)	1,223 (332)

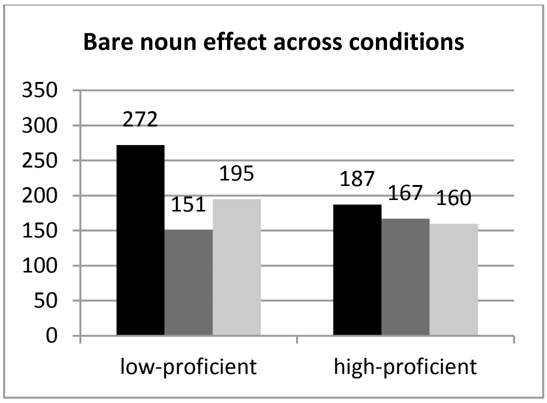


Table 4.21 RTs and standard deviations (in parentheses) of the monolingual group and low- and high-proficient bilingual groups for pooled cognates and noncognates (overall), for noncognates and cognates for bare noun naming across Gender Compatibility conditions.

Figure 4.8 The obtained bare noun effect for pooled cognates and noncognates (overall), for noncognates and cognates across Gender Compatibility conditions in the two proficiency groups of the bilingual group. (Bare noun effect = mean RTs of bilingual group minus mean RTs of monolingual group).

In the low-proficient group, statistical analyses revealed that the interaction of Gender Compatibility (only bare nouns) with Experiment Group was significant for F_1 and marginally significant for F_2 in the overall analysis ($F_1(2, 60) = 3.327$, $p = 0.043$; $F_2(2, 106) = 2.858$ $p = 0.062$). Follow-up analyses revealed that the effect of Gender Compatibility (only bare nouns) was significant in the F_1-analysis and marginally significant in the F_2-analysis for the monolingual group ($F_1(2, 38) = 7.140$, $p = 0.005$; $F_2(2, 106) = 2.661$, $p = 0.075$) but not significant in any of the analyses for the bilingual group ($F_1(2, 22) = 1.792$, $p = 0.190$; $F_2(2, 106) = 0.694$, $p = 0.502$). Furthermore, the interaction of Gender Compatibility (only bare nouns) with Experiment Group was not significant in the noncognate analysis ($F_1(2, 96) = 1.075$, $p = 0.345$; $F_2(2, 65) = 1.115$, $p = 0.334$) and neither in the cognate analysis ($F_1(2, 96) = 1.697$, $p = 0.189$; $F_2(2, 38) = 2.308$, $p = 0.113$).

In the high-proficient group, statistical analyses revealed that the interaction of Gender Compatibility (only bare nouns) with Experiment Group was not significant in the overall analysis ($F_1(2, 72) = 0.875$, $p = 0.421$; $F_2(2, 106) = 0.724$ $p = 0.487$), and neither in the noncognate analysis ($F_1(2, 72) = 0.999$, $p = 0.373$; $F_2(2, 65) = 0.567$, $p = 0.570$) nor the cognate analysis ($F_1(2, 72) = 0.438$, $p = 0.647$; $F_2(2, 38) = 0.151$, $p = 0.860$).

Summary

In the analyses of the NP interference effect, there were no significant interactions of Experiment Group with Phrase Type and Gender Compatibility, neither for low-proficient nor high-proficient bilinguals. At face value, the descriptive pattern of the NP effect for noncognates the low-proficient bilinguals was into the expected direction. However, for these items the interaction of Experiment Group with Phrase Type and Gender Compatibility only reached marginal significance in the F_2-analysis. There was a significant interaction of Experiment Group with Phrase Type in the analysis of both proficiency groups. Bilinguals groups had significantly longer naming times in the NP condition than the bare noun naming condition. This means that the bilingual groups had significantly more difficulties with determiner production than the monolingual group. Both analyses revealed a significant effect of Experiment Group as the monolingual group had faster naming times than the bilingual groups.

In the analyses of the bare noun interference effect, except for a significant effect in the F_1-analysis and a marginally significant effect in the F_2-analysis of the overall analysis of the low-proficient bilingual group, no significant interactions of Experiment Group with Gender Compatibility were found. Follow-up analyses revealed that the significant effect in the overall analysis of the low-proficient group stemmed from differences between Gender Compatibility conditions in the monolingual rather than the bilingual group.

4.3.3 Analysis of error rates

Bilingual German group

In the NP naming condition, determiner errors, or put differently, gender errors, were analyzed. Since contrary to the RT analysis no subjects had to be excluded due to missing observations, the analysis was based on 43 subjects. There were 21 high-proficient subjects and 22 low-proficient subjects. 21 of them had participated first in the LDT and 22 first in the PNT. Translation errors and unknown items as assessed in the translation task were removed per participant before analysis. No items were excluded so that the analysis was based on all 114 items.

		Correct	Incorrect	% incorrect
Low- proficient	Congruent	796	9	1.1 %
	Incongruent	751	26	3.3 %
	Incongruent n	790	27	3.3 %
High-proficient	Congruent	735	5	0.7 %
	Incongruent	707	5	0.7 %
	Incongruent n	723	13	1.8 %

Table 4.22 Amount of absolute correct and incorrect determiner productions, and percentage of incorrect productions, across Gender Compatibility conditions, for each proficiency group.

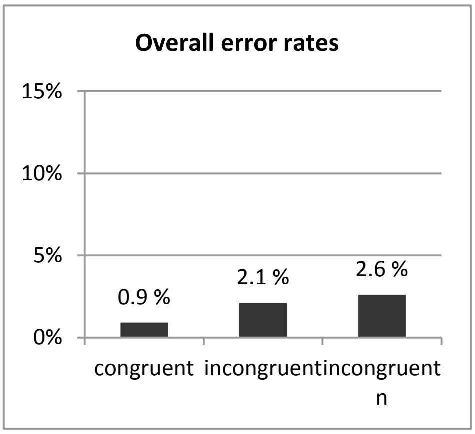


Figure 4.9 Overall percentage of error rates in determiner production per Gender Compatibility condition.

Looking at the descriptive data shown in Table 4.22 and Figure 4.9, it seems as if the tendency of article errors across Gender Compatibility conditions goes into the expected directions for both proficiency groups, that is, more errors are committed in the two incongruent conditions than in the congruent condition. However, it has to be noted that the overall error rate regarding determiners was very low: Overall, only 85 determiner errors were observed, which is less than one error per subject per condition.

Statistical analyses revealed that the differences across Gender Compatibility conditions were significant in the F_1-analysis ($F_1(2, 82) = 4.812$, $p = 0.015$, $F_2(2, 111) = 1.252$, $p = 0.290$). *T*-tests (all one-tailed[25] and Bonferroni-corrected) revealed that the difference between the congruent and incongruent condition was not significant ($t_1 = -1.761$, df = 42, $p = 0.129$; $t_2 = -1.627$, df = 74, $p = 0.162$), the difference between the congruent and incongruent neutral condition was only significant for t_1 ($t_1 = -3.355$, df = 42, $p = 0.003$; $t_2 = -1.338$, df = 74, $p = 0.282$) and the difference between the incongruent and incongruent neutral condition was not significant ($t_1 = -1.003$, df = 42, $p = 0.483$; $t_2 = -0.459$, df = 74, $p = 0.972$).

There was also a main effect of Level ($F_1(1, 41) = 7.429$, $p = 0.009$, $F_2(1, 111) = 15.746$, $p < 0.001$) and the interaction of Gender Compatibility with Level was significant in F_2 ($F_1(2, 82) = 2.136$, $p = 0.133$, $F_2(2, 111) = 3.152$ $p = 0.047$). Follow-up analyses of the interaction of Gender Compatibility with Level revealed that the difference across conditions was only significant in the F_1-analysis for both levels (low-proficient group: $F_1(2, 44) = 3.781$, $p = 0.043$, $F_2(2, 111) = 1.717$ $p = 0.184$; high-proficient group: $F_1(2, 38) = 4.022$, $p = 0.026$, $F_2(2, 111) = 1.176$ $p = 0.312$). *T*-tests (all one-tailed and Bonferroni-corrected) revealed that for the low-proficient group, the difference between the congruent and incongruent condition was marginally significant for t_2 ($t_1 = -2.017$, df = 22, $p = 0.168$; $t_2 = -2.357$, df = 74, $p = 0.063$). The difference between the congruent and incongruent neutral condition was significant for t_1 ($t_1 = -3.203$, df = 22, $p = 0.012$; $t_2 = -1.398$, df = 74, $p = 0.504$). The difference between the incongruent and incongruent neutral condition was not significant ($t_1 = 0.136$, df = 22, $p = 0.168$; $t_2 = 0.153$, df = 74, $p = 1.000$). For the high-proficient group, the difference between the congruent and incongruent condition was not significant ($t_1 = -0.112$, df = 19, $p = 1.000$; $t_2 = 0.000$, df = 74, $p = 1.000$) and neither was the difference between the congruent and incongruent neutral condition ($t_1 = -2.064$, df = 19, $p = 0.159$; $t_2 = -1.144$, df = 74, $p = 0.771$). The difference between the incongruent and incongruent neutral condition was significant for t_1 ($t_1 = -2.880$, df = 19, $p = 0.030$; $t_2 = -1.211$, df = 74, $p = 0.696$).

However, due to the low error rate (barely one error per subject per condition) also the significant results have to be interpreted with caution.

Bilingual Spanish group

The analysis was based on 38 subjects. There were 20 high-proficient subjects and 18 low-proficient subjects. 18 of them had participated first in the LDT and 20 first in the PNT. As before, translation errors and unknown items as assessed in the translation task were removed per participant before analysis. No items were excluded so that the analysis was based on all 114 items.

As can be seen in Table 4.23 and Figure 4.10, the amount of errors in determiner production differs across conditions, with the least amount of errors in the congruent condition and more errors committed in the two incongruent condition. This is according to the predictions (cf. section 4.1.2). The total number of determiner errors was 419.

[25] Here and in later places, one-tailed *t*-tests will be carried out whenever a hypothesis is clearly directional. In this case, e.g., the hypothesis was directional because more errors were expected for cognates than for noncognates, due to the lower frequency and lower transparency of the experimental cognate items.

		Correct	Incorrect	% incorrect
Low- proficient	**Congruent**	570	70	10.9 %
	Incongruent	512	83	13.9 %
	Incongruent n	550	89	13.9 %
High-proficient	**Congruent**	692	43	5.9 %
	Incongruent	635	51	7.4 %
	Incongruent n	641	83	11.5 %

Table 4.23 Amount of absolute correct and incorrect determiner productions, and percentage of incorrect productions, across Gender Compatibility conditions, for each proficiency group.

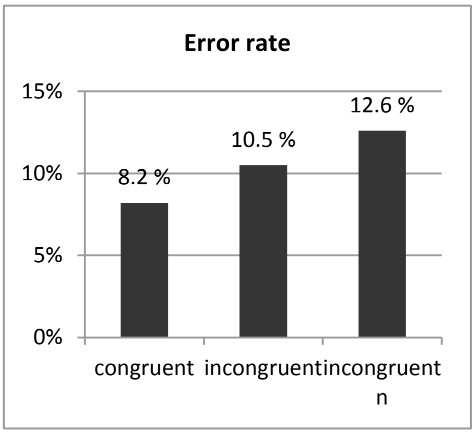


Figure 4.10 Overall percentage of error rates in determiner production per Gender Compatibility condition.

Statistical analyses revealed that the effect of Gender Compatibility was significant in the F_1-analysis ($F_1(2, 72) = 6.830$, $p = 0.004$, $F_2(2, 110) = 2.223$, $p = 0.113$). *T*-tests (all one-tailed and Bonferroni-corrected) revealed that the difference between the congruent and incongruent condition was marginally significant in the t_1-analysis ($t_1 = -2.448$, df = 37, $p = 0.057$; $t_2 = -0.792$, df = 74, $p = 1.000$), the difference between the congruent and incongruent neutral condition was significant for t_1 and marginally significant for t_2 ($t_1 = -3.106$, df = 37, $p = 0.012$; $t_2 = -2.208$, df = 74, $p = 0.090$) and the difference between the incongruent and incongruent neutral condition was not significant ($t_1 = -1.842$, df = 37, $p = 0.219$; $t_2 = -1.318$, df = 74, $p = 0.577$).

There was also a main effect of Level ($F_1(1, 36) = 8.133$, $p = 0.007$, $F_2(1, 110) = 32.125$, $p < 0.001$). The interaction of Gender Compatibility with Level was not significant ($F_1(2, 72) = 1.868$, $p = 0.170$, $F_2(2, 110) = 1.747$ $p = 0.179$).

Summary

When looking at the gender errors in the determiner naming condition, the error rates of both the German and the Spanish bilingual group exhibited differences across Gender Compatibility conditions. For both L2 groups, more errors were committed in the two incongruent conditions than in the congruent condition. Statistical analyses revealed that for the German bilingual group, the effect of Gender Compatibility was significant in the F_1-analysis. *T*-tests showed that the difference between the congruent and incongruent condition was marginally significant for t_1 and the difference between the congruent and incongruent neutral condition was significant for t_1 and marginally significant for t_2. The interaction of Gender Compatibility and Level was also significant in the F_2-analysis. Follow-up analyses revealed that the most important comparison between the congruent and incongruent condition was marginally significant for t_2 in the low-proficient group and not significant in the high-proficient group. This suggests that transfer occurred in the low-proficient group but not in the high-proficient group. However, due to the low amount of observations (barely one error per subject per condition on average), these results have to be interpreted with caution.

The analysis of the Spanish bilingual error rates revealed a significant effect of Gender Compatibility in the subject-analysis. *T*-tests showed that the difference between the congruent and incongruent condition was marginally significant in the t_1-analysis. The difference between the incongruent and incongruent neutral condition was significant for t_1 and marginally significant for t_2. There was a main effect of Level in both bilingual groups as more errors were committed by low-proficient subjects than high-proficient subjects.

4.4 Discussion PNT

In the RT analyses, no significant gender interference effects were found, neither in the analysis of NP naming nor in the analysis of bare noun naming. In general, the tendencies of RTs across Gender Compatibility conditions were not into the expected direction[26]. In addition, RT patterns were mostly inconsistent between Cognate Status conditions[27]. Hence, there were not even trends into the expected direction. Unfortunately, when looking at the results of the monolingual control groups, there were some significant effects of Gender Compatibility in F_1-analyses. This suggests that for some reason the naming difficulty of the pictures varied across conditions.

One possible explanation for the failure to find a gender interference effect in RTs and the inconsistent RT patterns across the relevant conditions in the bilingual group is that in each task, each subject named half of the pictures with a determiner and the other half without. That way, Phrase Type was not a complete within-subjects variable but the bare noun naming RTs of some subjects were compared to the NP naming RTs of the same items of other subjects. It is possible that the inter-subject variability regarding picture naming (across the different conditions) was too high for the design to be well-balanced. This can also explain the effects of Gender Compatibility in the control groups. Especially in L2 acquisition, individual differences have been shown to play a great role (Osterhout et al., 2006). Furthermore, this explanation is corroborated by the fact that standard deviations of bilinguals in the present PNT were much higher than in the experiment by Lemhöfer, Spalek, and Schriefers (2008). In the experiment by Lemhöfer et al. (2008), standard deviations were roughly between 100 and 200 ms (cf. table 6, p. 321), while in the present experiment standard deviations were approximately between 300 and 400 ms for both bilingual groups (cf. Table 4.12 and Table 4.17). Even for the monolingual groups, standard deviations were oftentimes higher than 300 ms (cf. Table 4.9 and Table 4.15). What possibly adds to the loss of balance in the design in the bilingual case is the exclusion of items due to translation errors and errors in the PNT which can give rise to differences between conditions despite careful item-matching. While care was taken to use simple items that would be known by low-proficient L2 speakers, this source of imbalance could not be completely prevented. Loss of items is difficult to avoid in bilingual experiments. This is especially true for the Spanish subjects who had very high error rates in the PNT. Furthermore, it seems problematic to conduct the LDT and PNT with the same subjects as task order effects were observed, possibly adding to the already high variance.

Nevertheless, it is interesting that there was a significant effect of Phrase Type in the Spanish bilingual group. Naming times in the NP naming condition were significantly longer than in bare noun naming. Apparently, bilingual Spanish subjects experienced difficulties when they were required to produce the correct determiner. This is in line with the results by Lemhöfer et al. (2008, p. 320), who also found significantly longer naming times in NP naming than bare noun naming for native German subjects in their L2 Dutch. For the bilingual German subjects in the present experiment, however, there was no significant difference between naming times with or without determiner. The reason for this will be considered in the overall discussion of this experiment. Moreover, the Phrase Type

[26] Except for the NP effect in the low-proficient Spanish bilingual group.
[27] Except for the bare noun effect in the low-proficient Spanish bilingual group

effect in the Spanish bilingual group was found for both proficiency levels, which means that this difficulty to name pictures with determiner persisted even at higher proficiency levels.

As mentioned above, the high error rates in the PNT of especially the Spanish bilingual group yielded the possibility to analyze the error rates for effects of gender congruency. Because of the high data loss in bilingual experiments it is not unlikely to find effects in the error rates rather than in the RTs. This was also the case in the PNT conducted by Lemhöfer et al. (2008; cf. section 3.3.2). Also in the present experiment, some effects of Gender Compatibility emerged in the error rates, which were consistent with the predictions (cf. section 4.1.2). The error rates of the German bilingual group showed an effect into the expected direction, with more errors committed in the incongruent conditions than in the congruent condition. However, so few determiner errors were committed (on average hardly one error per participant per condition) that this result has to be interpreted with caution. For the same reason, the interaction of Gender Compatibility with Level (significant for F_2) is probably spurious or at least not very robust. The Spanish bilingual group also made more errors in the two incongruent conditions than in the congruent condition. They committed more errors overall so that this result seems more robust. The main effect of Gender Compatibility was significant in the F_1-analysis. Post-hoc analyses revealed that the difference between the congruent and incongruent condition was marginally significant in the t_1-analysis, the difference between the congruent and incongruent neutral condition was significant for t_1 and marginally significant for t_2. Thus, in the Spanish bilingual group, there was clear tendency towards an effect of Gender Compatibility on error rates.

To summarize, in the present PNT, there was no gender interference effect in RTs but tendencies towards a gender transfer effect were visible in the error rates for both bilingual groups. This is in line with the results of the PNT conducted by Lemhöfer et al. (2008) who also only found a Gender Compatibility effect in the error rates but not in the RTs. Bordag and Pechmann (2007), on the other hand, only found gender congruency effects in the RTs but not in the error rates, where the difference between the congruent and incongruent condition failed to reach significance. The gender congruency effect in bare noun naming found by Morales, Paolieri, and Bajo, (2011) and Paolieri et al. (2010) could not be replicated.

Next, the Method and Results of the offline gender task will be discussed.

4.5 Offline gender assignment task

In the offline gender assignment task, subjects had to assign the correct determiner to the items also used in the PNT and LDT. Similar to the PNT or any language production task, for successful completion the correct determiner has to be retrieved so the same type of knowledge as in the PNT is tested. However, different from an online task such as the PNT, in the offline gender assignment task, subjects are not under time pressure but can take their time to think about the correct answer. This makes the task easier and it is possible that more determiners are assigned correctly than in an online task. The importance of time pressure and other task demands was discussed in section 2.2. Another advantage of an offline gender assignment task is that no items will be lost due to naming errors or other sources of errors typical for a PNT. Furthermore, the actual knowledge and the correctness of L2 gender representations can be assessed. An offline gender assignment task was also conducted by Lemhöfer, Schriefers, and Hanique (2010) and Lemhöfer, Spalek, and Schriefers (2008). In both studies, effects of gender compatibility were observed. If gender transfer occurs in offline gender assignment, more errors should be made in the two incongruent conditions than in the congruent condition.

4.5.1 Method offline gender assignment task

As mentioned before, the offline gender assignment task was always administered after completing the PNT. It was only filled out by the bilingual participants and for all the experimental items (in the L2). Items were presented in alphabetical order. Bilingual participants had to write down the correct article for each of the nouns, specify the certainty of their response on a scale ranging from 0 to 7 (0 = not sure at all, 7 = very sure), and indicate the familiarity of each word on the same scale (0 = not familiar at all, 7 = very familiar).

Data analysis

Similar to Lemhöfer, Spalek, and Schriefers (2008), GLMs with repeated measures were carried out on error percentages. Factors were Gender Compatibility and Level. In the F_1-analysis, Gender Compatibility was a within-subjects factor and Level a between-subjects factor. In the F_2-analysis, Level was a within-items factor and Gender Compatibility a between-items factor. Greenhouse-Geisser corrections (Greenhouse-Geisser, 1959) were applied in cases where sphericity could not be assumed. Reported degrees of freedom are uncorrected but p-values are corrected. In the analysis of German subjects with Spanish subjects together L1 was included as an additional factor.

4.5.2 Results offline gender assignment task

Analysis German subjects

The analysis was based on the data of the 44 subjects who entered data analysis of the LDT. Unknown, differently, and incorrectly translated words were removed (6.5 %, absolute count 326) before analysis. Table 4.24 shows the error rates of the low- and high-proficient group per Cognate Status and Gender Compatibility condition.

Overall error rates were quite low (2.8 %, absolute count 133), but more errors were made for cognates (5.2 %, 93) than for noncognates (1.4 %, 40). This difference showed a trend towards significance (U = 1,337.000, $N_1 = 72$, $N_2 = 42$, $p = 0.090$, one-tailed). Furthermore, except for the low-proficient group in the cognate condition, no big differences across Gender Compatibility conditions can be observed. Low-proficient subjects committed more errors (4.2 %, 101) than high-proficient subjects (1.4 %, 32) in all conditions.

		Low-proficient group			High-proficient group			Overall
		Correct	Incorrect	% incorrect	Correct	Incorrect	% incorrect	% incorrect
Overall	**Congruent**	775	30	3.7 %	763	12	1.5 %	2.7 %
	Incongruent	738	39	5.0 %	744	2	0.3 %	2.7 %
	Incongruent n	784	32	3.9 %	753	18	2.3 %	3.2 %
Noncognates	**Congruent**	474	15	3.1 %	483	4	0.8 %	1.9 %
	Incongruent	461	7	1.5 %	457	0	0 %	0.8 %
	Incongruent n	491	10	2.0 %	479	4	0.8 %	1.4 %
Cognates	**Congruent**	301	15	4.7 %	280	8	2.8 %	3.8 %
	Incongruent	277	32	10.4 %	287	2	0.7 %	5.7 %
	Incongruent n	293	22	7.0 %	274	14	4.9 %	6.0 %
Total mean		4.2 %			1.2 %			2.8 %

Table 4.24 Error rates (absolute counts of correct and incorrect assignments, percentage of incorrect assignments) of the German low- and high-proficient group per Gender Compatibility condition, for pooled cognates and noncognates (overall), and for noncognates and cognates separately.

Statistical analyses

A GLM with repeated measures (cf. Table 4.25) revealed, that there was a main effect of Level in the overall and noncognate analysis for F_1 and F_2 and for cognates only for F_1. The interaction of Gender Compatibility with Level was marginally significant in the F_1-analysis of the overall analysis and significant for both F_1 and F_2 in the cognate analysis. In order to further investigate the latter interaction effect, another analysis of the cognate items separate for each Level was carried out. The results of the analysis showed that there was no significant effect of Gender Compatibility within the cognate items for low-proficient subjects ($F_1(2, 44) = 2.245$, $p = 0.142$; $F_2(2, 39) = 0.515$, $p = 0.601$). For high-proficient subjects, there was only a significant effect in the F_1-analysis ($F_1(2, 40) = 7.477$, $p = 0.005$; F_2 $(2, 39) = 0.682$, $p = 0.512$). Considering the low error rate of high-proficient subjects in the cognate condition (2.8 %, absolute count 24), this result cannot be considered very robust. In addition, the tendency of the error rates across condition is not into the expected direction.

		F1			*F2*		
		df	***F***	***p***	***df***	***F***	***p***
Overall	**Gender Compatibility**	2, 84	.386	.633	2, 111	.044	.957
	Gender Compatibility * Level	2, 84	2.924	.072	2, 111	2.164	.120
	Level	1, 42	18.349	< .001*	1, 111	16.844	< .001*
Noncognates	**Gender Compatibility**	2, 84	1.854	.163	2, 69	.439	.646
	Gender Compatibility * Level	2, 84	.376	.687	2, 69	.198	.821
	Level	1, 42	12.638	.001*	1, 69	6.329	.014*
Cognates	**Gender Compatibility**	2, 84	1.251	.283	2, 39	12.013	.001*
	Gender Compatibility * Level	2, 84	4.445	.027*	2, 39	3.820	.031*
	Level	1, 42	12.382	.001*	1, 39	.133	.876

Table 4.25 Results of the F_1- and F_2-analyses, overall and for noncognates and cognates separately, with the factors Level and Gender Compatibility. Significant *p*-values ($p < 0.05$) are marked with an asterisk.

Summary

The most important finding is that there was no evidence for L1 gender transfer. Besides that, the overall error rate was very low (2.8 %). Apparently, German subjects had few problems with assigning the correct gender to Spanish nouns. Low-proficient German subjects made significantly more errors than high-proficient German subjects. The error rate for cognates (5.2 %) was more than three times higher than for noncognates (1.4 %). The higher error rate for cognates could be due to the fact that significantly more cognates than noncognates had intransparent endings (57 % vs. 14 %, respectively; $X^2 = 23.712$, df = 1, $p < 0.001$). Cognates also had a somewhat lower logarithmic frequency than noncognates (1.08, SD 0.43 vs. 1.33, SD 0.50), respectively). This difference was also significant ($t = 2.702$, df = 112, $p = 0.008$). The effect of transparency and frequency on the error rates of the German group as well as frequency effects on the error rates of the Spanish group will be further investigated below (cf. section Additional analyses).

Analysis Spanish subjects

The analysis was based on the data of the 39 subjects who entered data analysis of the LDT. Unknown, differently, and incorrectly translated items were removed (7.36 %, absolute count) before analysis. Table 4.26 shows the error rates of the low- and high-proficient group per Cognate Status and Gender Compatibility condition.

		Low-proficient group			High-proficient group			Overall
		Correct	Incorrect	% incorrect	Correct	Incorrect	% incorrect	% incorrect
Overall	Congruent	536	140	20.7 %	654	81	11.0 %	15.7 %
	Incongruent	428	195	31.3 %	544	142	20.7 %	25.7 %
	Incongruent n	451	224	33.2 %	543	181	25.0 %	28.9 %
Noncognates	Congruent	325	85	20.7 %	415	42	9.2 %	14.6 %
	Incongruent	267	110	29.2 %	341	85	20.0 %	24.3 %
	Incongruent n	294	118	28.6 %	343	102	22.9 %	25.7 %
Cognates	Congruent	211	55	20.7 %	239	39	14.0 %	17.3 %
	Incongruent	161	85	34.6 %	203	57	21.9 %	28.1 %
	Incongruent n	157	106	40.3 %	200	79	28.3 %	34.1 %
Total mean		28.3 %			18.8 %			23.4 %

Table 4.26 Error rates (absolute counts of correct and incorrect assignments, percentage of incorrect assignments) of the Spanish low- and high-proficient group per Gender Compatibility condition, for pooled cognates and noncognates (overall), and for noncognates and cognates separately.

Overall error rates were quite high (23.4 %, absolute count 963) but better than chance (binomial test: $p < 0.001$). More errors were made in the cognate (26.4 %, 421) than in the noncognate condition (21.4 %, 542). This difference was not significant (U= 1,410.500, $N_1 = 72$, $N_2 = 42$, $p = 0.275$, one-tailed). Furthermore, error rates differed across Gender Compatibility conditions, with the least errors committed in the congruent condition and more errors committed in the two incongruent conditions. Low-proficient subjects committed more errors (28.3 %, 559) than high-proficient subjects (18.8 %, 404) in all conditions.

Statistical Analyses

A GLM with repeated measures (cf. Table 4.27) revealed that there was a significant effect of Gender Compatibility in the overall analysis. In the noncognate and cognate analyses, this effect only reached significance in the F_1-Analyses. *T*-tests (all one-tailed) showed that the difference between the congruent and incongruent condition was significant (t_1 = -5.507, df = 38, $p < 0.001$; t_2 = -2.258, df = 74, $p = 0.007$), as well as the difference between the congruent and incongruent neutral condition (t_1 = -6.115, df = 38, $p < 0.001$; t_2 = -3.075, df = 74, $p = 0.002$). The difference between the two incongruent conditions was not significant (t_1 = -1.394, df = 38, $p = 0.086$; t_2 = -0.783, df = 74, p = 0.218). There was no significant interaction with Level but there was a significant main effect of Level in all three analyses.

		F1			F2		
		df	*F*	*p*	*df*	*F*	*p*
Overall	**Gender Compatibility**	2, 74	23.244	< .001*	2, 111	4.637	.012*
	Gender Compatibility * Level	2, 74	.161	.852	2, 111	.803	.451
	Level	1, 37	9.106	.005*	1, 111	18.349	< .001*
Noncognates	**Gender Compatibility**	2, 74	15.127	< .001*	2, 69	2.756	.071
	Gender Compatibility * Level	2, 74	.932	.398	2, 69	1.115	.334
	Level	1, 37	7.002	.012*	1, 69	24.747	< .001*
Cognates	**Gender Compatibility**	2, 74	15.089	< .001*	2, 39	1.840	.172
	Gender Compatibility * Level	2, 74	.532	.589	2, 39	1.037	.364
	Level	1, 37	10.062	.003*	1, 39	33.131	< .001*

Table 4.27 Results of the F_1- and F_2-analyses, overall and for noncognates and cognates separately, with the factors Level and Gender Compatibility. Significant *p*-values ($p < 0.05$) are marked with an asterisk**.**

Summary

Spanish subjects had fairly high error rates (23.4 %), much higher than the German subjects (2.8 %). Apparently, assigning the correct gender to German nouns was quite problematic for the Spanish subjects. Contrary to the German group, in the Spanish group, error rates differed significantly across pronoun conditions, with the lowest error rate in the congruent condition and higher error rates for the two incongruent conditions. As in the German group, high-proficient Spanish subjects made significantly less errors than low-proficient Spanish subjects. Spanish subjects also made slightly more errors in the cognate than in the noncognate condition, which was not significant.

Note that it is unlikely that the different error rates across Gender Compatibility conditions were caused by some difference between the conditions other than Gender Compatibility. Besides Gender Compatibility, item difficulty across conditions with respect to gender assignment should be similar as items were carefully matched (cf. section 4.2.2). Moreover, word familiarity ratings (indicated on a scale from 0 to 7; 0 = not familiar at all, 7 = very familiar) between conditions (congruent: 6.66, SD 0.88; incongruent: 6.67, SD 0.94; incongruent neutral 6.68, SD 0.87) did not differ significantly (all $p >$ 0.450).

Analysis German and Spanish subjects

In order to get a better understanding of the differences between the Spanish and German subject groups, especially the possible different influences of Gender Compatibility, another analysis with the additional factor L1 (German/Spanish) was carried out. A GLM with repeated measures with the factors L1, Level, and Gender Compatibility was conducted. As before, unknown, differently, and incorrectly translated words were not included in the analysis. Results are reported in Table 4.28.

	F1	F2

		df	F	p	df	F	p
Overall	Gender Compatibility	2, 158	24.255	< .001*	2, 111	3.876	.024*
	Gender Compatibility * L1	2, 158	21.684	< .001*	2, 111	4.286	.016*
	L1	1, 79	194.530	< .001*	1, 111	116.741	< .001*
	L1 * Level	1, 79	16.511	< .001*	1, 111	62.390	< .001*
	Level	1, 79	4.740	.032*	1, 111	23.556	< .001*
Noncognates	Gender Compatibility	2, 158	13.359	< .001*	2, 69	2.360	.102
	Gender Compatibility * L1	1, 158	18.169	< .001*	2, 69	2.980	.057
	L1	1, 79	170.692	< .001*	1, 69	82.851	< .001*
	L1 * Level	1, 79	11.129	< .001*	1, 69	29.222	< .001*
	Level	1, 79	4.894	.030*	1, 69	16.397	< .001*
Cognates	Gender Compatibility	2, 158	17.224	< .001*	2, 39	1.591	.217
	Gender Compatibility * L1	1, 158	10.179	< .001*	2, 39	1.353	.270
	L1	1, 79	165.462	< .001*	1, 39	34.093	< .001*
	L1 * Level	1, 79	19.562	< .001*	1, 39	39.137	< .001*
	Level	1, 79	3.060	.084	1, 39	7.602	.009*

Table 4.28 Results of the F_1- and F_2-analyses, overall and for noncognates and cognates separately, with the factors L1, Level and Gender Compatibility. Significant *p*-values ($p < 0.05$) are marked with an asterisk. Only significant effects are displayed.

First of all, there was a main effect of L1 in all three analyses because Spanish subjects made significantly more errors than German subjects (23.4 % vs. 2.8 %, respectively). There was also a main effect of Level in all three analyses, as error rates for high-proficient subjects (9.8 %) were significantly lower than for low-proficient subjects (15.1 %). The interaction of L1 and Level was also a significant.

Furthermore, there was a main effect of Gender Compatibility in the overall analysis. For the noncognate and cognate analysis, this effect was only significant in the F_1-analyses. This effect was primarily driven by the Spanish subject group: The likewise significant interaction of L1 and Gender Compatibility confirms the results of the two previous analyses, where a main effect of Gender Compatibility for the Spanish group but not for the German group was found.

Summary

The analysis pooled over German and Spanish subjects revealed that the Spanish subjects made significantly more errors in the offline gender assignment task than the German subjects. The analysis also confirmed that the effect of Gender Compatibility was different depending on the L1. The separate analyses of German and Spanish subjects had revealed that there was an effect of Gender Compatibility for the Spanish group but not the German group.

These findings are also reflected in the results of the subjective gender certainty ratings. Gender certainty (indicated on a scale from 0 to 7; 0 = not sure at all, 7 = very sure) was quite high for both groups but a little higher for the German group (6.64, SD 1.13) than for the Spanish group (5.74, SD 1.75). This difference in gender certainty between the two groups was significant (U= 162.000, $N_1 = 39$, $N_2 = 44$, $p < 0.001$). Apparently, the difference in accuracy between the two subject groups is reflected in their gender certainty ratings. Not surprisingly, gender certainty was also significantly higher for high-proficient subjects than for low-proficient subjects in both L1 groups. For German high-proficient subjects, mean gender certainty was 6.84 (SD 0.62) and 6.45 (SD 1.43) for low-proficient subjects (U = 126.000, $N_1 = 23$, $N_2 = 21$, $p = 0.004$, one-tailed). For Spanish high-proficient subjects, mean gender certainty was 5.85 (SD 1.71) and 5.61 (SD 1.79) for low-proficient subjects (U= 161.000, $N_1 = 19$, $N_2 = 20$, $p = 0.208$, one-tailed).

Additional analyses

As we saw before in the analysis of German subjects and in the analysis of Spanish subjects, German as well as Spanish subjects made more gender assignment errors in the cognate condition than in the noncognate condition. This difference was marginally significant for German subjects but not for Spanish subjects. As mentioned before, cognate items in general were less frequent than noncognates and Spanish cognates had less transparent nouns endings than Spanish noncognates. Therefore, I wanted to investigate if error rates for cognates and overall error rates in the German group were (more strongly) affected by the transparency of noun endings or word frequency. The effect of word frequency is also examined for the Spanish group. Another aim was to gain more insights on why overall error rates for Germans were so much lower than for Spanish subjects. To this end, differences in gender certainty will also be discussed.

Transparency and frequency effects German group

In the German group, the overall error rate for intransparent items was 7 % and for transparent[28] items 1 %[29]. This difference was significant (U = 842.000, $N_1 = 34$, $N_2 = 80$, $p < 0.001$). Thus, transparency had a significant influence on error rates. This difference in accuracy is also reflected in subjective certainty ratings (indicated on a scale from 0 to 7): Mean gender certainty for intransparent items was 6.38 (SD 1.37) and 6.76 (SD 0.98) for transparent items. This difference was also significant (U = 467.000, $N_1 = 34$, $N_2 = 80$, $p < 0.001$). Transparency of Spanish noun endings seemed to play an important role when assigning gender[30]. Frequency, on the other hand, did not correlate significantly with error rate ($\rho = -0.085$, N = 114, $p = 0.184$, one-tailed).

Apparently, in the German group, error rates were significantly lower for transparent noun endings, while frequency had no significant effect. Moreover, for German subjects, the error rate for

[28] As mentioned in section 4.2.2, in the present experiment only nouns ending in "*–o*" or "*–a*" were considered transparent as this is the simplest and most salient rule which can be assumed to be known even by beginning speakers of Spanish. All other items were considered intransparent.

[29] Note that as transparency was not a variable that was in the center of interest of the present study, the number of transparent and intransparent items was balanced across conditions but not across the whole experiment. 29.8 % (34) items had an intransparent noun ending, while 70.2 % (80) had a transparent noun ending.

[30] Transparency conditions were not confounded with frequency (BuscaPalabras corpus; $t = 0.0072$, df = 112, $p = 0.943$).

unknown[31] words was only 3.6 % (5 of 135). For unknown intransparent words the error rate was 20 % (3 out of 12) but only 1.6 % (2 out of 123) for unknown transparent words. This difference was significant ($U = 31.000$, $N_1 = 5$, $N_2 = 28$, $p = .001$, one-tailed). However, subjective certainty ratings for unknown transparent words (mean 5.66, SD 1.81) were still significantly lower than for transparent words subjects knew[32] (mean 6.76, SD 0.98; $U = 467.000$, $N_1 = 80$, $N_2 = 28$, $p = .001$).

Frequency effects Spanish group

The correlation between frequency and error rate was small but significant ($\rho = -0.159$, $N = 114$, $p = 0.045$, one-tailed). So the more frequent an item was, the lower the error rate. As discussed before, Spanish subjects also made more errors in the cognate condition (26.4 %, absolute count 421) than in the noncognate condition (21.4 %, 542), just like the German group, even though this difference was not significant. Nevertheless, the somewhat higher error rate for cognates could be caused by the lower frequency of the cognate items in this experiment[33].

Moreover, when only the unknown words were analyzed, in the Spanish group the error rate augmented to 44.7 % (42 out of 94, which is still significantly better than chance[34], as revealed by a binomial test: $p < 0.001$). This is very different from the German group whose error rate for unknown words (3.6 %) was only slightly higher than for known words (2.8 %). When looking at gender certainty, the same picture is seen. The mean gender certainty for unknown words for German subjects was quite high (5.48, SD 1.95) but quite low for Spanish subjects (2.88, SD 2.23). This difference in gender certainty for unknown words was significant ($U = 87.500$, $N_1 = 24$, $N_2 = 23$, $p < 0.001$).

Summary

It can be concluded that the error rates of the German group are strongly affected by transparency, whereas frequency does not seem to have much effect. The higher error rates for cognates can thus probably be attributed to a lack of transparent noun endings for the cognate items in the experiment. For Spanish nouns with transparent endings, the gender of infrequent and even unknown words can be easily inferred. So it is possible that frequency plays a bigger role when assigning gender to intransparent nouns. In future gender assignment experiments, it would be interesting to specifically manipulate item transparency and frequency to learn more about how these two factors interact. Moreover, in contrast to the German group, Spanish subjects were strongly affected by frequency. This is probably due to the general intransparency of German nouns and supports the assumption that frequency in gender assignment is important in the case of intransparent nouns. Furthermore, Spanish subjects behaved at chance when assigning articles to unknown words. Their certainty of gender assignments to unknown words was also significantly lower than for German subjects.

---

[31] For this analysis and similar subsequent analyses, only items where *no* translation equivalent was provided, i.e., the slot was left empty, were labeled “unknown translations” (cf. section 4.2.3 on error coding of the word translation task).

[32] Items that were incorrectly or differently translated were excluded from this analysis (n = 160).

[33] Note that this higher error rate is not caused by greater transfer effects in the case of form-similar words as in previous analyses no greater Gender Compatibility effect for cognates than for noncognates became apparent.

[34] Note that since German has a three-way gender system, the odds for making an incorrect guess are 2:3, i.e., 66 %.

4.5.3 Discussion offline gender assignment task

In the offline gender assignment task, German subjects performed almost at ceiling (error rate 2.8 %) and high-proficient subjects performed significantly better than low-proficient subjects. Contrary to the offline gender assignment task of Lemhöfer, Schriefers, and Hanique (2010) and Lemhöfer, Spalek, and Schriefers (2008) (both L1 German–L2 Dutch), no effect of Gender Compatibility on accuracy was found. For the Spanish group, quite a different pattern emerged. Spanish subjects not only made significantly more errors (23.4 %) than German subjects. They also showed substantial L1 gender transfer in the error rates, which is in line with the findings of Lemhöfer et al. (2010, 2008). However, contrary to these studies, in the present task this effect of Gender Compatibility did not seem to be mediated by Cognate Status. Gender transfer was also not mediated by proficiency, but high-proficient subjects performed significantly better than low-proficient subjects.

Interestingly, for German subjects, the error rate was especially low for transparent items (1 %) and significantly higher for intransparent items (7 %). Hence, German subjects´ overall very low error rate can be explained by the fact that the Spanish gender system is very simple and transparent. Therefore, the lack of a gender transfer effect for L1 Germans in the present experiment could also be due to a ceiling effect. Even in an analysis conducted with words that subjects could not translate, the error rate was still low (3.6 %). For the Spanish group, on the other hand, the overall high error rate was probably caused by the relative intransparency of the German gender system and the fact that German has a three-way gender system. In an analysis carried out with words that subjects did not know, the error rate was even higher (44.7 %). This is consistent with the finding that item frequency was correlated with accuracy for the Spanish group but not for the German group. This is probably due to the fact that item transparency played a greater role than frequency when assigning gender to Spanish nouns.

Hence, it is possible that frequency is more important in the case of intransparent item endings or, put more generally, in the case of intransparent gender systems, than when transparent gender cues are available. This assumption is supported by the fact that Sabourin, Stowe, and De Haan (2006) found frequency effects (or as they say, ”effects of noun familiarity” (p. 1)) in a gender assignment task for L2 Dutch, which also has a fairly intransparent gender system just as German. Unfortunately, though, due to the low number of intransparent Spanish items in the present experiment and the overall low error rate of the German subjects the interaction of transparency and frequency could not be tested in a reliable manner. Hence, future research is necessary to shed light on the combined effects of transparency and frequency in the learning of L2 gender. Nevertheless, the present results combined with Sabourin et al.´s findings suggest that frequency effects in L2 gender assignment are mediated by transparency.

This assumption is in line with DeKeyser´s (2005) discussion of factors that make form–meaning mapping difficult for late L2 learners. According to DeKeyser, frequency is especially important in the case of intransparent mappings, which are problematic for L2 learners. This is consistent with the frequency effect found for Spanish subjects in L2 German. When mapping is transparent, however, the structure can be acquired fast, as demonstrated by the low error rate of low-proficient German subjects in the present experiment (4.2 %). Furthermore, according to DeKeyser, in the case of intransparent mapping, late L2 learners will most certainly have problems and might even not be able to acquire the structure at all (fossilization). This is supported by the fact that even the high-proficient Spanish subjects continued to make many gender errors (18.8 %).

As mentioned above, Sabourin et al. investigated familiarity effects in the gender assignment task by investigating frequency effects (p. 8). They argued that "[…] the higher the frequency of a noun, the more experience and familiarity a speaker should have with it, […]". This is probably true, but nevertheless it seems a bit peculiar and unnecessary to equalize the concept of frequency with the concept of familiarity. In the present study I collected subjective familiarity ratings in addition to the frequency data of the items which allows me to examine the relationship between frequency and subjective familiarity and whether they are really interchangeable concepts. I obtained a significant correlation between logarithmic frequency and familiarity for both subject groups (German subjects: $\rho = 0.523$, $n = 144$, $p < 0.001$; Spanish subjects: $\rho = 0.560$, $n = 114$, $p < 0.001$, all one-tailed), however, the correlation was not very high. Thus, even if the two measures are correlated, because of the medium size correlation it is not justified to treat familiarity and frequency as completely interchangeable concepts. Familiarity is a subjective measure and frequency a more objective measure based on corpora which are probably more valid for L1 than L2 speakers. In the end, both frequency and familiarity are only indirect ways of assessing the amount of exposure bilinguals had to certain words and it is difficult to determine which of the two is the better operationalization. In each case, in the present experiment, the correlation between word familiarity and error rate was greater than for frequency and error rate for Spanish subjects[35] ($\rho = -0.374$, $N = 114$, $p < 0.001$; $\rho = -0.159$, $N = 114$, $p = 0.045$, both one-tailed, respectively). Assuming that more exposure leads to greater gender assignment accuracy, subjective familiarity seems to be a better predictor of accuracy and could possibly be a better operationalization for exposure than frequency of L1 corpora. Further research is necessary to determine which approach best approximates actual L2 exposure. Until then, it seems fruitful to collect data on various related variables.

Besides effects of frequency and transparency, also effects of gender certainty were investigated in the present experiment. In general, gender certainty was negatively correlated with error rate (Germans: $\rho = -0.457$, $N = 114$, $p < 0.001$; Spaniards: $\rho = -0.678$, $N = 114$, $p < 0.001$, both one-tailed), that is, the higher gender certainty was, the less errors were made. Hence, by and large, both L1 groups were not bad at estimating their L2 gender knowledge. However, when calculating the mean gender certainty across Gender Compatibility conditions for the two subject groups, I obtained the following results: Gender certainty ratings across Gender Compatibility conditions were not significantly different, neither for the German nor for the Spanish subjects (Kruskall-Wallis-test: $X^2 = 1.375$, $df = 2$, $p = 0.503$ and $X^2 = 0.651$, $df = 2$, $p = 0.722$, respectively). This is different from the result of Lemhöfer, Schriefers, and Hanique (2010) who also obtained a significant effect of Gender Compatibility on gender certainty ratings, with lower certainty ratings for the incongruent condition (p. 152). Apparently, in the present experiment, even when assigning gender in the two incongruent conditions, Spanish subjects were quite certain of their response (both means > 5.4 out of 7) even though overall they were – slightly, but significantly – less certain than German subjects. Thus, for the Spanish subjects, L1 transfer seems to be quite strong and persistent, even leading to high certainty scores in incongruent conditions.

Furthermore, gender certainty ratings were positively correlated with logarithmic frequencies (German subjects: $\rho = 0.313$, $N = 114$, $p < 0.001$; Spanish subjects: $\rho = 0.435$, $N = 114$, $p < 0.001$, both one-tailed). And the correlations between gender certainty and the more subjective measure of word familiarity for German and for Spanish subjects were still higher ($\rho = 0.384$, $N = 114$, $p < 0.001$

[35] Since as discussed above German subjects´ accuracy was primarily influenced by item transparency the correlation between error rate and familiarity was not calculated.

and ρ = 0.758, N = 114, *p* < 0.001, respectively), especially for the Spanish group. Again, the smaller correlation coefficients for the German group underline the lower importance of familiarity and frequency effects in L2 Spanish gender assignment in comparison to transparency. Moreover, as we have seen, the accuracy of Spanish subjects was affected by item frequency and there was a strong effect of Level on the accuracy of assignments for both L1 groups. Lemhöfer et al. (2010) claimed that "just passively receiving correct input from the L2 environment is not sufficient for changing incorrect gender representations" (p. 121). The present data suggest that exposure does seem to have at least some beneficial effect. Furthermore, the positive correlation between word familiarity and gender certainty confirms that especially for the Spanish group the more familiar words get with exposure, the more certain learners are of their gender.

In conclusion, because of the effects of noun transparency found for L2 Spanish, the present experiment showed that transfer can also be mediated by L2 characteristics, while so far studies had mostly looked at effects of L1 characteristics (e.g., Sabourin & Stowe, 2008; Sabourin et al., 2006; Sagarra & Herschensohn, 2011; cf. chapter 3). The present finding is in line with the results of the earlier described study investigating gender congruency effects by Bordag and Pechmann (2007, cf. section 3.3.2). They found effects of noun ending transparency on RTs and error rates in picture naming (experiment 3) and a GJT (experiment 4). No transparency effects were found by Salamoura and Williams (2007) but, as noted by the authors, the number of transparent items might have been too low to reveal any effects. Regarding my differential findings for L2 Spanish and L2 German, the following explanation seems plausible: "[…] one important factor might be the opaque nature of the gender system in the second language: In absence of reliable form-related cues (i.e., word endings as in Spanish or Italian) for word gender (like in German or Dutch), the learner tends to use L1 gender information […]. By contrast, when easy-to-learn rules govern the assignment of grammatical gender, L2 influences might be overruled, or might not even arise in the first place." (Lemhöfer, Spalek, & Schriefers, 2008, pp. 327-328). Furthermore, the fact that gender transfer effects were obtained for L1 German subjects by Lemhöfer et al. (2010) and Lemhöfer et al. (2008) shows that L1 speakers of German are able to experience gender transfer from their native language. Therefore, the results obtained here are not due L1 characteristics but L2 characteristics. Nevertheless, the present result does not preclude that transfer from German to Spanish might occur with more intransparent and infrequent Spanish nouns or under circumstances of greater time pressure or greater task demands, such as having to compute greater agreement distances (cf. section 2.2). This potential trade-off between L2 gender system transparency or simplicity, task demands, and proficiency is addressed in Experiment 2.

I will now discuss the Method and results of the LDT, investigating possible gender transfer effects in a comprehension task.

4.6 Introduction LDT

The present LDT aims to expand on the study by Lemhöfer, Spalek, and Schriefers (2008), who, to my knowledge, were the first to investigate gender interference in a visual word recognition task. As summarized in section 3.3.2, bilingual gender interference effects have usually been investigated in production tasks[36], and more specifically, in PNTs. However, just because a bilingual gender interference effect has by now oftentimes been found in language production, this does not

[36] Except for a few studies investigating gender transfer effects in sentence processing which are discussed in the introduction of Experiment 2 (cf. section 5.1).

automatically imply that gender interference also occurs in word recognition. As stated by Costa and Santesteban (2004) "[…] the nature of the processes involved in each are different enough to warrant caution in exporting assumptions from one modality to the other without independent motivation." (p. 253).

One crucial difference between word production and word comprehension studies is the direction of the activation flow. In language production, the information flow is top-down, while in comprehension it is bottom-up. This means that in word production, the flow starts at the conceptual level with the formulation of the preverbal message and ends with the phonological output. In word comprehension, the direction of activation is reversed. Here, first the orthographic/phonological information is perceived and finally leads to activation of the corresponding concept. This has important consequences for grammatical gender processing and might, more specifically, have different implications for languages with transparent and intransparent gender systems. In languages with a transparent gender system like Spanish, gender information can be directly activated from the orthographic noun ending. In languages with a more intransparent gender system like German, on the other hand, first the meaning (semantic level) has to be activated for the syntactic information to become available at the lemma level. In word production, the gender information has to be actively retrieved, while in word comprehension, this information can be passively activated. Therefore, it is relevant for models of bilingual language production and comprehension whether the direction of the activation flow has consequences for the gender interference effect and whether the same gender interference effect observed in language production can also be observed in comprehension. Furthermore, in language comprehension a cue to the target language is contained in the stimulus itself, while in language production the target language has to be selected by the speaker. This might affect the level of activation of the target and non-target language and has thus an impact on potential interference processes. Once again, Costa and Santesteban (2004) point out that "[…] the issue of the simultaneous activation of the two lexicons of a bilingual might have different answers in each modality […]." (p. 253). As discussed at length in section 3.3.2, bilingual gender interference effects in language production have mostly been studied with PNTs using NP and bare noun naming.

Also in the monolingual domain, gender priming effects have been studied more extensively in the area of language production than in language comprehension. In the few studies investigating gender priming in (reading) comprehension, effects for valid gender primes are not always obtained, while results showing that invalid primes inhibit processing of the subsequent target noun are more consistent (e.g., Gurjanov, Lukatela, Lukatela, Savic, & Turvey, 1985; Jacobsen, 1999; Jakubowicz & Faussart, 1998), as explained by Lemhöfer, Spalek, and Schriefers (2008). Following the reasoning of Lemhöfer et al. (2008), in the bilingual situation, L2 determiner primes of nouns that are gender-incongruent across the two languages might be comparable to invalid gender primes in a monolingual situation. That way, this "hidden incongruent" (p. 314) condition should inhibit processing of the target. Since, as mentioned before, inhibition effects caused by incongruent primes are reliably found in the monolingual literature, it might be possible to obtain similar results with primes that are gender-incongruent across languages.

In bilingual comprehension, so far, besides Lemhöfer, Spalek, and Schriefers (2008), few studies have investigated gender-primed processing. In two earlier mentioned studies (cf. section 2.1) using auditory comprehension it was shown that contrary to native speakers, L2 learners were not able to use gender primes to speed up processing. In an auditory LDT, Scherag, Demuth, Rösler, Neville, and

Röder (2004) showed that native English speakers could not take advantage of gender-congruent adjective primes in their L2 German relative to gender-incongruent adjective primes. Native German speakers, on the other hand, exhibited facilitation and inhibition effects in response to gender-congruent and incongruent adjective primes. In a similar fashion, Guillelmon and Grosjean (2001) demonstrated that, contrary to monolinguals, English–French late bilinguals were not influenced by gender-congruent and incongruent adjective primes when asked to repeat the noun of an auditorily presented NP. However, as explained in sections 3.2.2 and 3.3.1, the ability to use L2 morphosyntactic information in general and gender information in particular might also depend on L1 characteristics, such as if the L1 has or lacks gender. English lacks grammatical gender, therefore possibly rendering it more difficult for these bilingual subjects to use this type of syntactic information in L2 processing. So even if native English speakers are not able to use L2 gender cues effectively, native speakers of gendered languages might be able to do so. As explained in section 3.3.2, Lemhöfer et al. (2008) were able to show inhibition effects in bilinguals, who were native speakers of a gendered language, in a determiner-primed LDT. Consequently, a primed LDT is an appropriate tool to investigate the following question: Can subjects use the gender information of the prime to distinguish words from nonwords faster or are they influenced by the L1 gender information?

If gender interference occurs, RTs to the primed incongruent conditions should be slower than to the primed congruent condition, relative to the unprimed condition. For the bare noun interference effect, RTs in the congruent condition should in principle also be faster than in the incongruent conditions, relative to the monolingual control group. However, since the bare noun interference effect has so far not been investigated in an LDT, the effect might also be different from picture naming.

Notably, there is an important parallel between the two comprehension experiments reported in this thesis, the present LDT and Experiment 2. In both the LDT and in Experiment 2, gender agreement has to be processed: in the LDT between determiner prime and noun and in Experiment 2 between anaphor and referent.

4.7 Method LDT

4.7.1 Material LDT

The same Spanish and German words as in the PNT were used (cf. section 4.2.2). In addition to the experimental material described earlier, also nonwords and filler items had to be created for the LDT.

Nonwords: Nonwords were obtained by changing one letter of an existing word. They were created from the pool of words that was left over after the matching process and were therefore similar to the experimental items. Care was taken for the nonwords to be approximately equally long in letters and syllables as the real words. Characteristics of nonwords compared to words can be seen in Table 4.29. In order to avoid any imbalances in the experiment, the amount of nonwords created from cognates and noncognates and from congruent, incongruent, and incongruent neutral words was equal.

Filler items: In order to avoid imbalances regarding grammatical gender in the experiment and to make up for the unequal distribution of gender in the cognate items, namely a lack of Spanish feminine words in the two incongruent conditions, 24 filler items (8 feminine in Spanish and masculine in German, 10 feminine in Spanish and neutral in German, 6 feminine in Spanish and in German) were included.

			Mean (SD)	Range	Mean (SD)	Range
Spanish	# of syllables	Nonword	2.50 (.605)	2 - 4	2.64 (.821)	1 - 4
		Word	2.57 (.624)	1 - 4	2.86 (.926)	1 - 5
	# of letters	Nonword	5.83 (1.151)	3 - 8	6.21 (1.646)	4 - 10
		Word	6.06 (1.481)	3 - 9	6.50 (2.063)	3 - 11
German	# of syllables	Nonword	1.71 (.458)	1 - 2	2.31 (1.568)	1 - 4
		Word	1.51 (.503)	1 - 2	2.55 (1.675)	1 - 4
	# of letters	Nonword	5.39 (1.302)	3 - 8	6.07 (.680)	3 - 10
		Word	5.10 (1.275)	2 - 8	6.21 (.803)	4 - 10

Table 4.29 Overview of the number of syllables and number of letters of Spanish and German nonwords and words per Cognate Status condition. Means, standard deviations (in parentheses), and the range are displayed.

In the end, the following overall distribution/selection of 252 items (50 % nonwords) was obtained: 84 cognates (words and nonwords), 144 noncognates (words and nonwords), 24 filler items (words and nonwords). There were 131 Spanish masculine items and 121 feminine items. Of the German items, 83 were masculine, 83 feminine, and 86 neutral. Additionally, at the beginning of each experimental block two warming-up items which were not analyzed were presented. Two of the four warming-up items had German feminine gender, one German masculine gender, and one German neutral gender. The Spanish gender of two items was feminine and masculine for the two other items. 16 additional items (8 words and 8 nonwords) were used as practice items with gender, gender compatibility, cognate status, and word type counterbalanced as in the experiment (cf. section 4.2.2).

An overview of the number of syllables and letters of the Spanish and German words and nonwords is given in Table 4.29.

4.7.2 Procedure LDT

If the LDT was their first session, participants did the DIALANG test to assess their language competence prior to the experiment in order to assure that subjects had the required level to participate.

Next, participants read the instructions and questions were answered by the instructor. The bilinguals received the instruction in the L2 in order to put them in their bilingual language mode (Grosjean, 1999). A short practice phase followed so that participants could get used to the procedure. After completing the LDT, participants did the word translation task and finally, participants filled in the language history questionnaire.

Words were presented in 28 upper case Geneva, in the center of the screen. Response keys for lexical decisions were the right shift key for words and the left shift key for nonwords. Half of the words appeared shortly after a definite determiner, half without a preceding determiner, depending on the prime list to which the participant was assigned (cf. section 4.2.2). Each participant saw each item only once, either with or without a determiner prime. The LDT was divided into two blocks between which participants could take a break as long as they wished. Each block started with two warming-up items which were not analyzed. It took about 15 - 20 minutes to complete the LDT.

Each trial started with a fixation cross that was presented for 700 ms. After a blank screen for 100 ms, either the determiner prime or a “placeholder” for the Spanish/German articles appeared about 50 pixels to the left of the center. Identical to the cues used for the NP condition in the PNT, the placeholders consisted of two low dashes for the Spanish version and three low dashes for the German version. After 250 ms the noun was presented about 50 pixels to the right of the center while the determiner prime or placeholder stayed on screen. Stimulus and prime/placeholder remained until the participant responded or till timeout after 3000 ms. The intertrial interval was 1000 ms.

4.8 Results LDT

In the following sections I will report the results of the LDT in Spanish (German bilingual and Spanish monolingual participants) and in German (Spanish bilingual and German monolingual participants). For each language, I will first look at the results of the monolingual control groups in order to verify whether items were correctly matched and to check for possible artifacts caused by the experimental design. As in the PNT, I will then summarize the results of the bilingual experimental group. In the end, I will discuss the analysis of the bilingual group together with the monolingual control group, also similar to the PNT. Data analyses are conducted in the same way as for the PNT and as described in section 4.2.4. Different from the PNT, error rates were not analyzed because overall error rates were very low[37].

4.8.1 Results LDT in Spanish

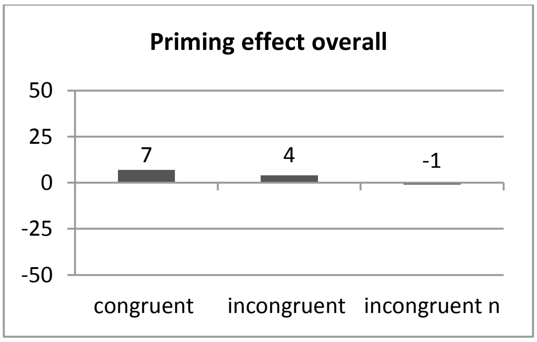
Results Spanish monolingual control group

The analysis was based on 20 subjects. There were 3.3 % (absolute count 75) wrong key presses and no failures to respond. These erroneous responses to the LDT were removed from the analysis. There were no outliers.

An overview of the RTs across Gender Compatibility and Phrase Type conditions and of the priming effect is given in Table 4.30 and Figure 4.11.

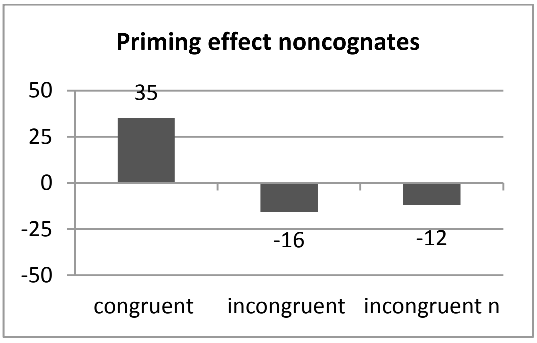
Overall

	Without prime	With prime
Congruent	674 (246)	681 (239)
Incongruent	684 (221)	688 (239)
Incongruent n	691 (255)	690 (269)



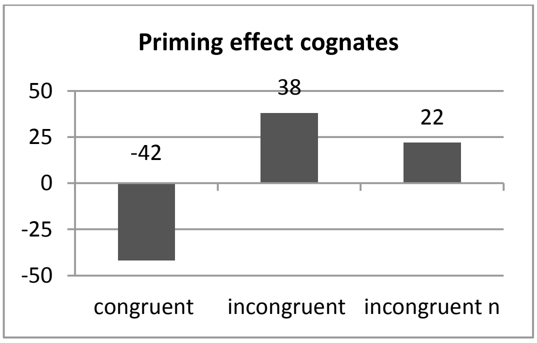
Noncognates

	Without prime	With prime
Congruent	663 (234)	698 (257)
Incongruent	676 (217)	660 (199)
Incongruent n	668 (220)	656 (197)



Cognates

	Without prime	With prime
Congruent	694 (264)	652 (202)
Incongruent	698 (226)	736 (290)
Incongruent n	729 (303)	751 (354)



[37] For both language groups, overall error rates (wrong key presses and failures to respond) in the LDT were lower than the amount of determiner errors committed in the LDT. German bilinguals pressed the wrong key 1.7 % (80) of the times in the LDT and failed to respond 0.4 % (21) of the times but made 5.3 % (85) determiner errors in the PNT. Spanish bilinguals pressed the wrong key 3.1 % (134) of the times in the LDT and failed to respond 0.2 % (7) of the times but made 9.3 % (318) determiner errors in the PNT. Note that in addition, determiner errors are much more informative regarding gender transfer processes than overall error rates which contain a lot of noise due to other processes hampering lexical decisions.

Table 4.30 RTs and standard deviations (in parentheses) for pooled cognates and noncognates (overall), for noncognates and cognates across Gender Compatibility and Prime Type conditions.

Figure 4.11 The obtained priming effect for pooled cognates and noncognates (overall), for noncognates and cognates across Gender Compatibility conditions. (Priming effect = mean RTs of with prime condition minus mean RTs of without prime condition).

Statistical analyses

The results of the three analyses can be seen in Table 4.31. There were no significant effects in any of the three analyses (overall, noncognates, cognates) except for a significant main effect of Gender Compatibility in the cognate analysis, which was only significant in the F_1-analysis.

		F1			*F2*		
		df	*F*	*p*	*df*	*F*	*p*
Overall	**Gender Compatibility**	2, 38	.976	.386	2, 111	.254	.776
	Prime Type	1, 19	.063	.804	1, 111	.094	.759
	Gender Compatibility * Prime Type	2, 38	.271	.764	2, 111	.062	.940
Noncognates	**Gender Compatibility**	2, 38	1.147	.328	2, 69	.367	.694
	Prime Type	1, 19	.014	.909	1, 69	.000	.997
	Gender Compatibility * Prime Type	2, 38	2.399	.104	2, 69	2.629	.079
Cognates	**Gender Compatibility**	2, 38	8.329	.001*	2, 39	1.536	.228
	Prime Type	1, 19	.009	.925	1, 39	.185	.669
	Gender Compatibility * Prime Type	2, 38	2.096	.137	2, 39	1.822	.175

Table 4.31 Results of the F_1- and F_2-analyses, overall, for noncognates and cognates, with the factors Gender Compatibility and Prime Type. Significant results are marked with an asterisk.

Item-matching

In order to assess if item-matching was successful, RTs for the different conditions without prime were examined (Table 4.30, columns "without prime"). ANOVAs revealed that the differences across conditions without prime were not significant, neither overall ($F_1(2, 38) = 1.302$, $p = 0.284$; $F_2(2, 111) = 0.307$, $p = 0.736$), nor for noncognates ($F_1(2, 38) = 1.610$, $p = 0.213$; $F_2(2, 69) = 0.116$, $p = 0.891$), nor for cognates ($F_1(2, 38) = 1.610$, $p = 0.213$; $F_2(2, 39) = 0.414$, $p = 0.664$).

Summary

Except for a significant effect of Gender Compatibility in the cognate F_1-analysis, there were no significant main or interaction effects of Gender Compatibility or Prime Type in the control group. Since the relevant interactions of Gender Compatibility with Phrase Type did not reach significance, the analyses of the bilingual group were conducted as planned. Regarding item-matching, statistical analysis revealed that unprimed RTs across Gender Compatibility conditions did not differ significantly from each other so items seemed to be well-matched.

Results bilingual German group

The analysis was based on 44 subjects. There were 21 high-proficient subjects and 23 low-proficient subjects. Half of them had participated first in the LDT and the other half first in the PNT. Translation errors and unknown items as assessed in the translation task were removed per participant before analysis. The overall error rate in the LDT was very low, there were 1.7 % wrong key presses (absolute count 80) and 0.4 % (21) failures to respond. These erroneous responses were also removed. No outliers were observed. Furthermore, three items (*florero–Vase, fusil–Gewehr, camello–Kamel*) had to be excluded from the analysis due to a lack of observations in some conditions of the item analysis. 111 items were left in the analysis, 70 (out of 72) noncognates and 41 (out of 42) cognates. In each of the Gender Compatibility conditions, the following amount of items was left: congruent condition = 38 items, incongruent condition = 37 items, incongruent neutral condition = 36 items. For noncognates, there were 24 items left in the congruent condition, 23 in the incongruent condition and 23 in the incongruent neutral condition. For cognates, there were 14 items in the congruent condition, 14 in the incongruent condition and 13 in the incongruent neutral condition.

An overview of the RTs is given in Table 4.32[38].

			PNT 1st			PNT 2nd		
			Without prime	**With prime**	**Prime Type effect**	**Without prime**	**With prime**	**Prime Type effect**
Low-proficient group	**Noncognates**	**Congruent**	784 (282)	822 (294)	+ 38	847 (254)	861 (268)	+ 14
		Incongruent	830 (282)	803 (246)	- 27	827 (209)	871 (281)	+ 44
		Incongruent n	788 (264)	772 (211)	- 16	856 (262)	833 (228)	- 23
	Cognates	**Congruent**	890 (465)	845 (390)	- 45	854 (319)	849 (261)	- 5
		Incongruent	810 (324)	842 (308)	+ 32	882 (328)	816 (197)	- 66
		Incongruent n	845 (330)	950 (483)	+ 105	894 (372)	904 (291)	+ 10
			PNT 1st			**PNT 2nd**		
			Without prime	**With prime**	**Prime Type effect**	**Without prime**	**With prime**	**Prime Type effect**
High-proficient group	**Noncognates**	**Congruent**	847 (248)	889 (274)	+ 42	741 (207)	761 (254)	+ 20
		Incongruent	875 (263)	878 (288)	+ 3	730 (204)	743 (191)	+ 13
		Incongruent n	840 (281)	869 (262)	+ 29	732 (234)	755 (206)	+ 23
	Cognates	**Congruent**	957 (429)	920 (436)	- 37	731 (225)	747 (248)	+ 16
		Incongruent	863 (289)	882 (262)	+ 19	760 (233)	748 (212)	- 12
		Incongruent n	1.049 (509)	1.011 (441)	- 38	758 (230)	853 (318)	+ 95

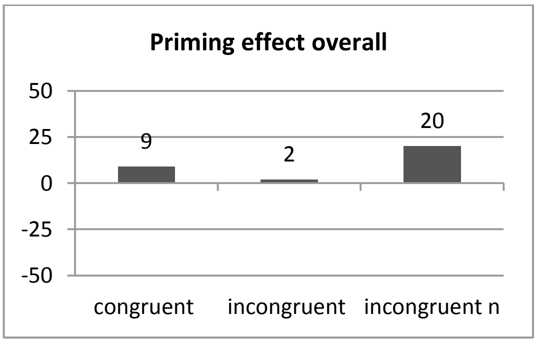
Table 4.32 Overview of the RTs and standard deviations (in parentheses) per Level (low- vs. high-proficient group), Cognate Status (noncognates vs. cognates), Gender Compatibility condition (congruent, incongruent, and incongruent neutral condition), Task Order (LDT first vs. LDT second), and Prime Type (without prime vs. with prime). Prime Type effect = mean RTs of with prime condition minus mean RTs of without prime condition.

As described in section 4.1.2, I expected either a small facilitation or no effect for the primed congruent condition and significant inhibition for the primed incongruent conditions, compared to the unprimed conditions. As can be seen in Figure 4.12, the observed effects looked different and were inconsistent across Cognate Status conditions.

[38] Note that when looking at Table 4.32 it could appear that the priming effects across conditions for cognates of the low-proficient group were into the expected direction when the LDT was the first task. The priming effect was - 45 ms for the congruent condition, + 32 for the incongruent condition and + 105 for the incongruent neutral condition. In a separate analysis, this effect was investigated. However, the interaction of Gender Compatibility with Prime Type was not significant ($F1(2,22) = 1.932$, $p = 0.169$; $F2(2,38) = 1.153$, $p = 0.119$) for the low-proficient group when the LDT was the first task and no other significant effects were found.

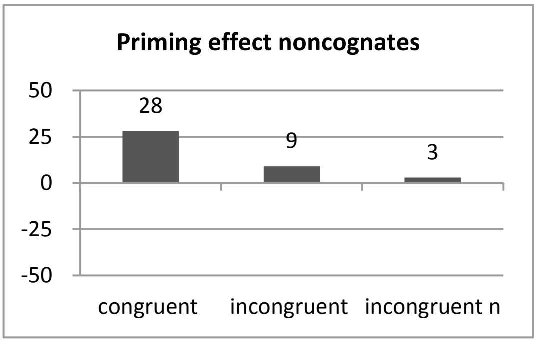
Overall

	Without prime	With prime
Congruent	824 (307)	833 (303)
Incongruent	819 (268)	821 (256)
Incongruent n	830 (312)	850 (307)



Noncognates

	Without prime	With prime
Congruent	803 (252)	831 (276)
Incongruent	813 (245)	822 (258)
Incongruent n	802 (264)	805 (231)



Cognates

	Without prime	With prime
Congruent	856 (377)	838 (344)
Incongruent	827 (298)	820 (252)
Incongruent n	879 (379)	927 (394)

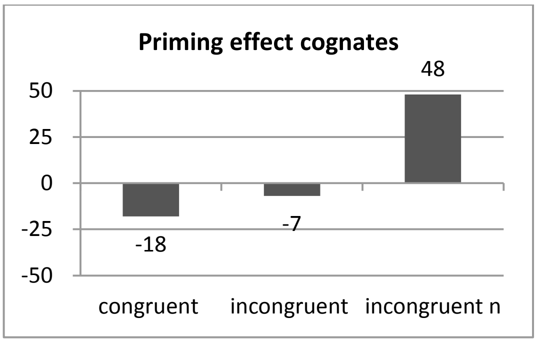


Table 4.33 RTs and standard deviations (in parentheses) for pooled cognates and noncognates (overall), for noncognates and cognates across Gender Compatibility and Prime Type conditions.

Figure 4.12 The obtained priming effect for pooled cognates and noncognates (overall), for noncognates and cognates across Gender Compatibility conditions. (Priming effect = mean RTs of with prime condition minus mean RTs of without prime condition).

As shown in Table 4.34, the interaction of Gender Compatibility with Prime Type was not significant and there were no significant higher-order interaction including this interaction in any of the three analyses. In the following, the data and effects are described in more detail.

		F1			*F2*		
		df	***F***	***p***	***df***	***F***	***p***
Overall	**Gender Compatibility**	2, 80	2.540	.085	2, 108	.294	.746
	Prime Type	1, 40	.747	.393	1, 108	1.306	.256
	Gender Compatibility * Prime Type	2, 80	.489	.615	2, 108	.462	.631
	Level	1, 40	.224	.638	1, 108	6.796	.010*
	Task Order	1, 40	1.912	.174	1, 108	57.823	< .001*
	Level * Task Order	1, 40	4.075	.050	1, 108	139.907	< .001*
Noncognates	**Gender Compatibility**	2, 80	1.029	.362	2, 67	.255	.776
	Prime Type	1, 40	1.573	.217	1, 67	1.272	.263
	Gender Compatibility * Prime Type	2, 80	1.369	.260	2, 67	1.234	.298
	Level	1, 40	.432	.515	1, 67	8.128	.006*
	Task Order	1, 40	.956	.334	1, 67	28.800	< .001*
	Level * Task Order	1, 40	4.505	.040*	1, 67	72.065	< .001*
Cognates	**Gender Compatibility**	2, 80	13.747	< .001*	2, 38	1.134	.332
	Gender Compatibility * Task Order	2, 80	2.708	.073	2, 38	1.855	.170
	Prime Type	1, 40	.135	.715	1, 38	.168	.684
	Gender Compatibility * Prime Type	2, 80	1.946	.150	2, 38	2.324	.112
	Level	1, 40	.019	.892	1, 38	.668	.419
	Task Order	1, 40	3.400	.073	1, 38	38.022	< .001*
	Level * Task Order	1, 40	3.386	.073	1, 38	74.141	< .001*

Table 4.34 Results of the overall, noncognate, and cognate F_1- and F_2-analysis with the factors Gender Compatibility, Prime Type, Level, and Task Order. Effects are only displayed if they are (a) theoretically important (i.e., main effects and interaction effect of Gender Compatibility and Prime Type), (b) if their *p*-value is < .10. Main effects and interactions that are not relevant with regard to the predictions and with a *p*-value > .10 are not displayed.

Overall analysis: The results are shown in Table 4.34. On the whole, determiner primes had little effect on RTs. Primes slowed down lexical decisions by only 11 ms on average (without prime: 824 ms, SD 296 ms; with prime: 835 ms, SD 290 ms), which was not significant. High-proficient subjects had faster RTs (819 ms, SD 288 ms) than low-proficient subjects (840 ms, SD 297 ms), though this difference was quite small (only 21 ms) and only significant in the F_2- analysis[39]. Regarding task order

[39] Note that nevertheless the difference between the two Levels (high- and low-proficient) was significant regarding the CEFR levels (χ^2 =44.000, df = 4, p < 0.001).

effects, subjects performed on average 55 ms faster if the LDT was the second task (LDT first: 858 ms, SD 325 ms, LDT second: 803 ms, SD 256 ms). This difference was only significant in the F_2-analysis. Moreover, there was also a significant interaction of Level and Task Order (i.e., significant for F_2 and almost significant in the F_1-analysis). RTs across Gender Compatibility conditions were largely similar, except for the incongruent neutral condition which had slightly longer RTs (congruent: 829 ms, SD 305 ms; incongruent: 820 ms, SD 262 ms; incongruent neutral: 840 ms, SD 310 ms). The effect of Gender Compatibility was marginally significant in the F_1-analysis.

Noncognate analysis (cf. Table 4.34): As in the overall analysis, there was a significant main effect of Level in the F_2-analysis, a significant main effect of Task Order in the F_2-analysis and an interaction effect of Level with Task Order in both the F_1- and F_2-analysis.

Cognate Analysis (cf. Table 4.34): As in the overall analysis, there was a significant main effect of Task Order in the F_2-analysis which was also marginally significant in the F_1-analysis. There was an interaction effect of Level with Task Order which was marginally significant in the F_1-analysis and significant in the F_2-analysis. There was also a significant effect of Gender Compatibility.

Summary

There was no interaction effect of Gender Compatibility with Prime Type in any of the three analyses. Moreover, RT patterns across Gender Compatibility conditions were different from the expected pattern and inconsistent between Cognate Status conditions.

Analysis of bilingual German group with monolingual Spanish control group

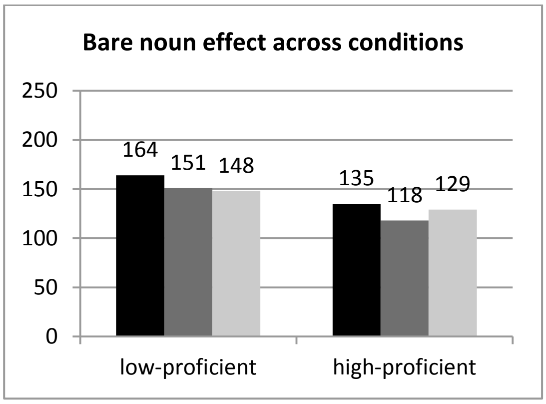
This analysis was conducted in order to investigate the bare noun gender interference effect. It is based on the same participants and items as the previous analysis.

Bare noun effect

The descriptive data of the bare noun effect shown in Table 4.35 and Figure 4.13 do not seem to reveal a consistent Gender Compatibility effect across the three analyses. None of the RT patterns corresponds to the expected bare noun effect, that is, faster RTs for the congruent condition than for the incongruent conditions.

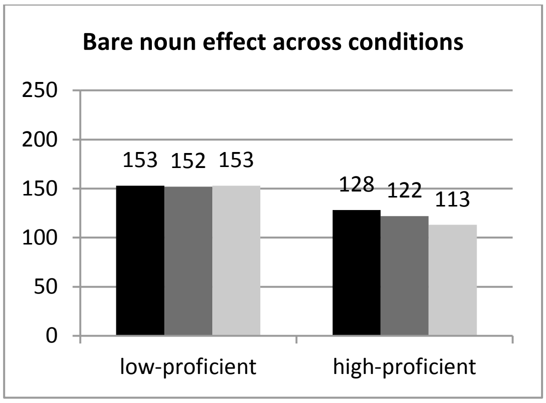
Overall

	Mono-linguals	Bilinguals Low	Bilinguals High
Congruent	674 (246)	838 (327)	809 (285)
Incongruent	684 (221)	835 (282)	802 (252)
Incongruent n	691 (255)	839 (299)	820 (326)



Noncognates

	Mono-linguals	Bilinguals Low	Bilinguals High
Congruent	663 (234)	816 (270)	791 (233)
Incongruent	676 (217)	828 (247)	798 (244)
Incongruent n	668 (220)	821 (264)	781 (262)



Cognates

	Mono-linguals	Bilinguals Low	Bilinguals High
Congruent	694 (264)	872 (398)	839 (355)
Incongruent	698 (226)	844 (326)	808 (265)
Incongruent n	729 (303)	869 (351)	890 (408)

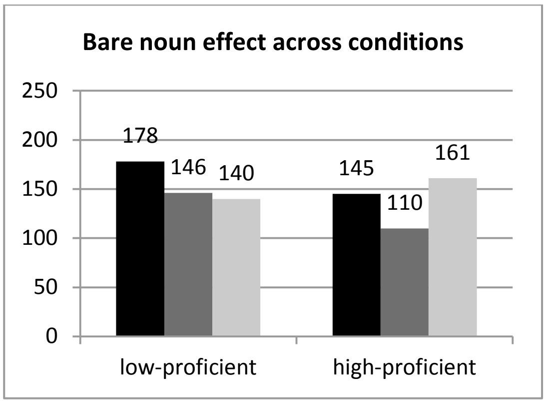


Table 4.35 RTs and standard deviations (in parentheses) of the monolingual and bilingual group for pooled cognates and noncognates (overall), for noncognates and cognates for bare noun processing across Gender Compatibility conditions.

Figure 4.13 The obtained bare noun effect for pooled cognates and noncognates (overall), for noncognates and cognates across Gender Compatibility conditions in the two proficiency groups of the bilingual group. (Bare noun effect = mean RTs of bilingual group minus mean RTs of monolingual group).

For the low-proficient group, statistical analyses revealed that the interaction of Gender Compatibility (only bare nouns) with Experiment Group was not significant, neither in the overall analysis ($F_1(2, 82) = 0.094$, $p = 0.911$; $F_2(2, 108) = 0.021$, $p = 0.979$), nor in the noncognate analysis ($F_1(2, 82) = 0.146$, $p = 0.864$; $F_2(2, 67) = 0.210$, $p = 0.811$), nor cognate analysis ($F_1(2, 82) = 0.571$, $p = 0.543$; $F_2(2, 38) = 0.230$, $p = 0.796$).

For the high-proficient group, statistical analyses revealed that the interaction of Gender Compatibility (only bare nouns) with Experiment Group was not significant, neither in the overall analysis ($F_1(2, 78) = 0.910$, $p = 0.407$; $F_2(2, 108) = 0.433$, $p = 0.650$), nor in the noncognate analysis ($F_1(2, 78) = 0.021$, $p = 0.979$; $F_2(2, 67) = 0.015$, $p = 0.985$), nor cognate analysis ($F_1(2, 78) = 0.983$, $p = 0.379$; $F_2(2, 38) = 0.695$, $p = 0.505$).

Summary

There were no significant interaction effects of Gender Compatibility with Experiment Group and thus no evidence for a bare noun gender interference effect.

4.8.2 Results LDT in German

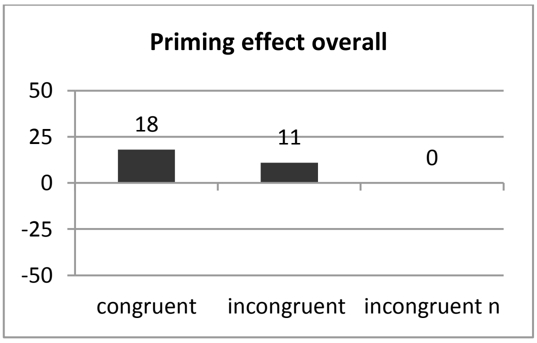
Results German monolingual control group

The analysis was based on 20 subjects. There were 1.67 % (absolute count 38) wrong key presses and 0.09 % (2) failures to respond. These erroneous responses to the LDT were removed from the analysis. There were no outliers.

An overview of the RTs across Gender Compatibility and Prime Type conditions and of the priming effect is given in Table 4.36 and Figure 4.1.

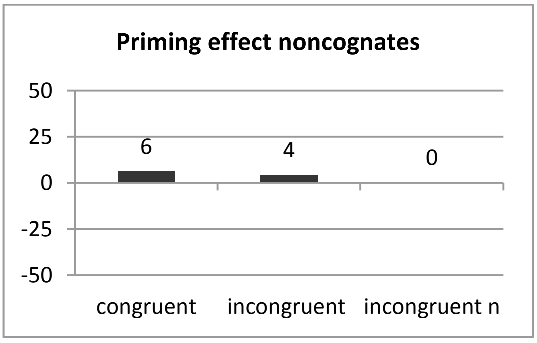
Overall

	Without prime	With prime
Congruent	680 (199)	698 (213)
Incongruent	701 (215)	712 (224)
Incongruent n	717 (228)	717 (248)



Noncognates

	Without prime	With prime
Congruent	655 (178)	661 (166)
Incongruent	680 (183)	684 (207)
Incongruent n	685 (186)	685 (189)



Cognates

	Without prime	With prime
Congruent	725 (226)	762 (264)
Incongruent	736 (257)	760 (245)
Incongruent n	773 (278)	775 (319)

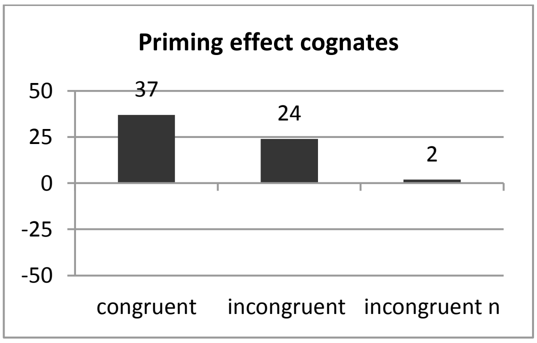


Table 4.36 RTs and standard deviations (in parentheses) for pooled cognates and noncognates (overall), for noncognates and cognates across Gender Compatibility and Prime Type conditions.

Figure 4.14 The obtained priming effect for pooled cognates and noncognates (overall), for noncognates and cognates across Gender Compatibility conditions. (Priming effect = mean RTs of with prime condition minus mean RTs of without prime condition).

Statistical analyses

The results of the three analyses can be seen in Table 4.37. There were no significant effects except for a significant main effect of Gender Compatibility in the overall and noncognate F_1-analysis which was not significant in the F_2-analysis.

		F1			***F2***		
		df	***F***	***p***	***df***	***F***	***p***
Overall	**Gender Compatibility**	2, 38	4.376	.019*	1, 111	1.199	.276
	Prime Type	1, 19	1.115	.304	2, 111	.250	.779
	Gender Compatibility * Prime Type	2, 38	.529	.594	2, 111	.689	.504
Noncognates	**Gender Compatibility**	2, 38	4.587	.016*	2, 69	1.188	.311
	Prime Type	1, 19	.202	.658	1, 69	.196	.660
	Gender Compatibility * Prime Type	2, 38	.050	.951	2, 69	.028	.972
Cognates	**Gender Compatibility**	2, 38	1.543	.227	2, 39	.209	.813
	Prime Type	1, 19	.931	.347	1, 39	1.112	.298
	Gender Compatibility * Prime Type	2, 38	.493	.614	2, 39	.266	.768

Table 4.37 Results of the F_1- and F_2-analyses, overall, for noncognates and cognates, with the factors Gender Compatibility and Prime Type. Significant results are marked with an asterisk.

Item-matching

RTs for the different conditions without prime were examined (Table 4.36, columns "without prime"). ANOVAs revealed that the differences across conditions without prime were significant in the F_1-analysis, overall for noncognates and cognates taken together ($F_1(2, 38) = 6.866$, $p = 0.003$; $F_2(2, 111) = 0.852$, $p = 0.429$), for noncognates ($F_1(2, 38) = 3.379$, $p = 0.045$; $F_2(2, 69) = 0.343$, $p = 0.355$), and for cognates ($F_1(2, 38) = 4{,}420$, $p = 0.032$; $F_2(2, 39) = 0.343$, $p = 0.712$).

Summary

Except for a significant effect of Gender Compatibility in the F_1-analysis of the overall and noncognate analysis there were no significant main or interaction effects of Gender Compatibility or Prime Type in the control group. Therefore, the analyses of the bilingual group were conducted as planned. Regarding item-matching, statistical analyses revealed that unprimed RTs across Gender Compatibility conditions unfortunately differed significantly from each other in the F_1-analysis of all three analyses. The reason for these significant differences in the F_1-analysis is unclear. Items were well-matched regarding the most important parameters such as frequency and number of letters (all $F < 1$; compare Method section). Only the number of syllables was slightly lower for the incongruent neutral condition (mean 1.38, SD 0.5) than for the incongruent and congruent condition (both 1.58, SD 0.5) in the case of noncognates ($F = 1.386$). However, it seems unlikely that this is the reason for the significant effects since RTs were higher, not lower, for the incongruent neutral condition (cf. Table 4.36, column "without prime"). So the origin of these differences between unprimed conditions in the control group remains unclear. The effects found here are especially surprising considering that there were no significant differences in the control group of the previously discussed Spanish version of the LDT even though the same items had been used and the same matching criteria had been applied. It is of course possible that the differences are due to an effect of an uncontrolled parameter such as bigram frequency or neighborhood density. Since, however, the focus of the present experiment was on an interaction between Prime Type and Gender Compatibility these differences in the F_1-analyses do not necessarily pose a problem.

Results bilingual Spanish group

The analysis was based on 39 subjects. There were 20 high-proficient subjects and 19 low-proficient subjects. 19 of them had participated first in the LDT and 20 first in the PNT. Translation errors and unknown items as assessed in the translation task were removed per participant before analysis. The overall error rate in the LDT was very low, there were 3.1 % wrong key presses (absolute count 134) and 0.2 % (7) failures to respond. These erroneous responses were also removed. No outliers were observed. Furthermore, four items (*nudo–Knoten, florero–Vase, disco–CD, computadora–computer*) had to be excluded from the analysis due to a lack of observations in some conditions of the subject and item analysis. 110 items were left in the analysis, 69 (out of 72) noncognates and 41 (out of 42) cognates. In each of the Gender Compatibility conditions, the following amount of items was left: congruent condition = 37 items, incongruent condition = 35 items, incongruent neutral condition = 38 items. For noncognates, there were 23 items left in the congruent condition, 22 in the incongruent condition and 24 in the incongruent neutral condition. For cognates, there were 14 items in the congruent condition, 13 in the incongruent condition and 14 in the incongruent neutral condition.

An overview of the RTs[40] is given in Table 4.38.

			PNT 1st			PNT 2nd		
			Without prime	**With prime**	**Prime Type effect**	**Without prime**	**With prime**	**Prime Type effect**
Low-proficient group	**Noncognates**	**Congruent**	899 (412)	915 (434)	+ 16	846 (260)	867 (304)	+ 21
		Incongruent	907 (424)	900 (319)	- 7	901 (342)	878 (314)	- 23
		Incongruent n	850 (287)	922 (406)	+ 72	829 (252)	830 (245)	+ 1

[40] When looking at Table 4.38, it could appear that the priming effects across conditions for cognates for the high-proficient group were into the expected direction when the LDT was the first task. The priming effect was - 9 ms for the congruent condition, + 43 ms for the incongruent condition, and + 24 ms for the incongruent neutral condition. In a separate analysis, this effect was investigated. However, the interaction of Gender Compatibility with Prime Type was not significant ($F_1(2,218) = 0.233$, $p = 0.794$; $F_2(2,38) = 0.372$, $p = 0.692$) and no other significant effects were found.

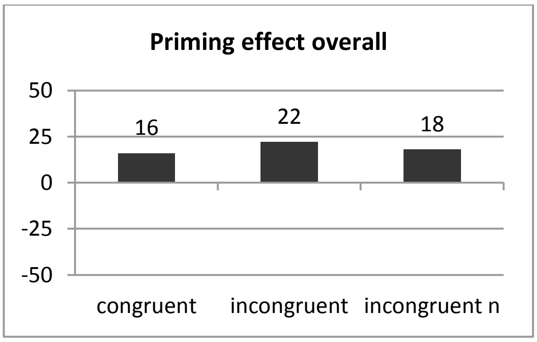
	Cognates	Congruent	939 (402)	1.012 (458)	+ 73	868 (390)	879 (326)	+ 11
		Incongruent	930 (436)	900 (317)	- 30	827 (337)	843 (282)	+ 16
		Incongruent n	1.053 (516)	997 (520)	- 56	879 (451)	884 (315)	+ 5
			PNT 1st			**PNT 2nd**		
			Without prime	**With prime**	**Prime Type effect**	**Without prime**	**With prime**	**Prime Type effect**
High-proficient group	**Noncognates**	**Congruent**	841 (348)	834 (287)	- 7	720 (221)	735 (185)	+ 15
		Incongruent	831 (307)	922 (409)	+ 91	733 (182)	753 (218)	+ 20
		Incongruent n	819 (321)	854 (310)	+ 35	694 (211)	751 (215)	+ 57
	Cognates	**Congruent**	884 (410)	875 (330)	- 9	722 (292)	761 (197)	+ 39
		Incongruent	906 (386)	949 (465)	+ 43	715 (181)	782 (350)	+ 67
		Incongruent n	951 (476)	975 (401)	+ 24	786 (309)	740 (179)	- 46

Table 4.38 Overview of the RTs and standard deviations (in parentheses) per Level (low- vs. high-proficient group), Cognate Status (noncognates vs. cognates), Gender Compatibility condition (congruent, incongruent, and incongruent neutral condition), Task Order (LDT first vs. LDT second), and Prime Type (without Prime vs. with Prime). Prime Type effect = mean RTs of with prime condition minus mean RTs of without prime condition.

As described in section 4.1.2, I expected either a small facilitation or no effect for the primed congruent condition and significant inhibition for the primed incongruent conditions, compared to the unprimed conditions. As can be seen in Figure 4.15, the observed effects looked different and were inconsistent across Cognate Status conditions.

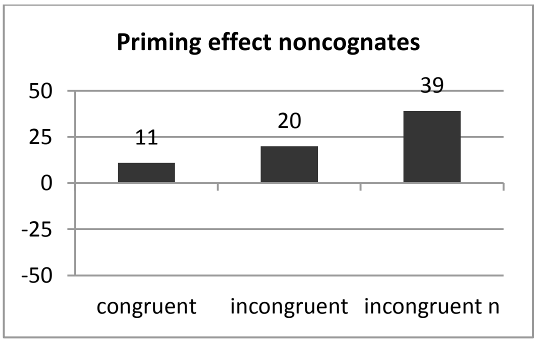
Overall

	Without prime	**With prime**
Congruent	834 (346)	850 (326)
Incongruent	841 (337)	863 (341)
Incongruent n	840 (355)	858 (335)



Noncognates

	Without prime	**With prime**
Congruent	822 (320)	833 (313)
Incongruent	840 (327)	860 (328)
Incongruent n	797 (277)	836 (304)



Cognates	Without prime	With prime
Congruent	851 (382)	877 (345)
Incongruent	842 (353)	866 (364)
Incongruent n	911 (449)	894 (377)

Table 4.39 RTs and standard deviations (in parentheses) for pooled cognates and noncognates (overall), for noncognates and cognates across Gender Compatibility and Prime Type conditions.

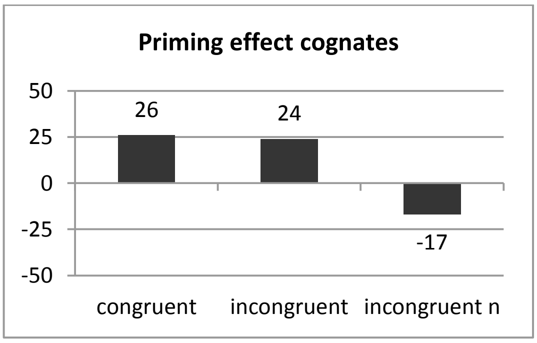


Figure 4.15 The obtained priming effect for pooled cognates and noncognates (overall), for noncognates and cognates across Gender Compatibility conditions. (Priming effect = mean RTs of with prime condition minus mean RTs of without prime condition).

Furthermore, as shown in Table 4.40, in the overall analysis the interaction of Gender Compatibility with Prime Type was not significant and there was no significant higher-order interaction including this interaction. Neither was there a significant interaction when the analysis was split up according to Cognate Status. In the following, the data and effects will be described in more detail.

Overall analysis: Results are displayed in Table 4.40. Lexical decisions were a little slower with determiner prime (857 ms, SD 334 ms) than without (838 ms, SD 346 ms). Furthermore, high-proficient subjects had much faster RTs (799 ms, SD 281 ms) than low-proficient subjects (900 ms, SD 388 ms), which was significant in the F_2-analysis and marginally significant in the F_1-analysis. Regarding task order effects, subjects performed on average 83 ms faster if the LDT was the second task (808 ms, SD 312 ms), compared to 891 ms if the LDT was the first task (SD 363 ms). This difference was significant in the F_1- as well as the F_2-analysis. Furthermore, the interaction of Level with Task Order was significant in the F_2-analysis but not the F_1-analysis.

		F1			***F2***		
		df	***F***	***p***	***df***	***F***	***p***
Overall	**Gender Compatibility**	2, 70	.338	.714	2, 107	.072	.931
	Gender Compatibility * Level	2, 70	1.135	.327	2, 107	2.396	.096
	Prime Type	1, 35	2.771	.105	1, 107	1.875	.174
	Gender Compatibility * Prime Type	2, 70	.072	.931	2, 107	.173	.841
	Level	1, 35	3.462	.071	1, 107	73.208	< .001*
	Task Order	1, 35	4.474	.042*	1, 107	83.054	< .001*
	Level * Task Order	1, 35	.556	.461	1, 107	12.664	.001*
Noncogn ates	**Gender Compatibility**	2, 70	1.907	.156	2, 66	.479	.622
	Prime Type	1, 35	3.418	.073	1, 66	2.928	.092

	Gender Compatibility * Prime Type	2, 70	1.159	.320	2, 66	.368	.693
	Level	1, 35	4.474	.042*	1, 66	51.220	< .001*
	Task Order	1, 35	3.138	.085	1, 66	36.324	< .001*
	Level * Task Order	1, 35	.839	.366	1, 66	8.663	.004*
Cognates	**Gender Compatibility**	2, 70	4.747	.012*	2, 38	.551	.581
	Gender Compatibility * Level	2, 70	2.558	.085	2, 38	2.621	.086
	Prime Type	1, 35	.238	.629	1, 38	.065	.800
	Gender Compatibility * Prime Type	2, 70	.791	.457	2, 38	.776	.467
	Level	1, 35	2.106	.156	1, 38	22.331	< .001*
	Task Order	1, 35	5.701	.022*	1, 38	53.708	< .001*
	Level * Task Order	1, 35	.352	.557	1, 38	3.724	.061

Table 4.40 Results of the overall, noncognate, and cognate F_1- and F_2-analysis with the factors Gender Compatibility, Prime Type, Level, and Task Order. Effects are only displayed if they are (a) theoretically important (i.e., main effects and interaction effect of Gender Compatibility and Prime Type), (b) if their *p*-value is < .10. Main effects and interactions that are not relevant with regard to the predictions and with a *p*-value > .10 are not displayed.

Noncognate analysis (cf. Table 4.40): There was a significant main effect of Level (significant in the F_1 and F_2-analysis) and a main effect of Task Order which was significant in the F_2-analysis and marginally significant in the F_1-analysis. The interaction of Level with Task Order was significant in the F_2-analysis but not the F_1-analysis.

Cognate analysis (cf. Table 4.40): As in the previous analyses, there was a main effect of Task Order which was significant in the F_1- and F_2-analysis. The effect of Level was only significant in the F_2-analysis. Furthermore, there was a significant effect of Gender Compatibility in the F_1-analysis.

Summary

Just as in the German bilingual group, there was no effect of Gender Compatibility with Prime Type and neither was there a significant interaction with these two factors. Furthermore, the RT patterns across Gender Compatibility conditions were different from the expected pattern and inconsistent between Cognate Status conditions so that also no tendencies into the expected direction became apparent.

Analysis of bilingual Spanish group with monolingual German control group

This analysis was conducted in order to investigate the bare noun gender interference effect. It is based on the same participants and items as the previous analysis.

Bare noun effect

The descriptive data of the bare noun effect shown in Table 4.41 and Figure 4.1 do not seem to reveal a consistent Gender Compatibility effect across the three analyses. None of the RT patterns corresponds to the expected bare noun effect, that is, faster RTs for the congruent condition than for the incongruent conditions.

Overall

	Mono-linguals	Bilinguals	
		Low	High
Congruent	680 (199)	884 (364)	787 (322)
Incongruent	701 (215)	892 (383)	793 (281)
Incongruent n	717 (228)	883 (368)	799 (338)

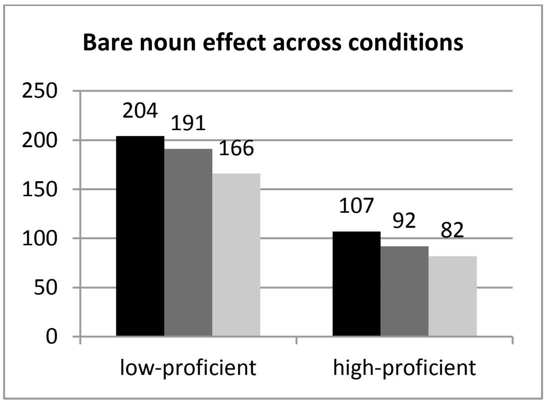
Noncognates

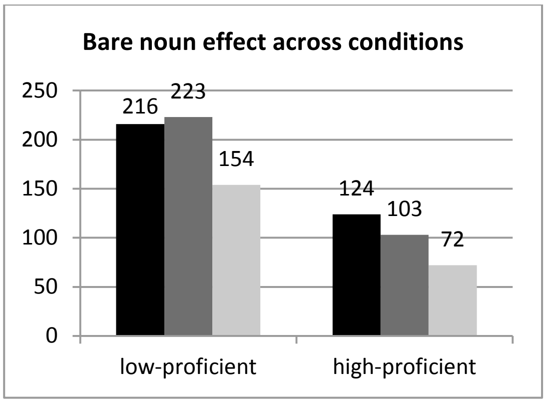
	Mono-linguals	Bilinguals	
		Low	High
Congruent	655 (178)	871 (341)	779 (295)
Incongruent	680 (183)	903 (381)	783 (258)
Incongruent n	685 (186)	839 (269)	757 (279)

Cognates

	Mono-linguals	Bilinguals	
		Low	High
Congruent	725 (226)	902 (396)	801 (362)
Incongruent	736 (257)	875 (388)	810 (315)
Incongruent n	773 (278)	958 (487)	869 (408)

Table 4.41 RTs and standard deviations (in parentheses) of the monolingual and bilingual group for pooled cognates and noncognates (overall), for noncognates and cognates for bare noun processing across Gender Compatibility conditions.





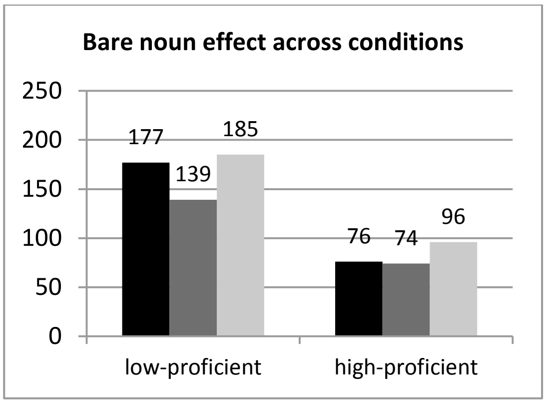


Figure 4.16 The obtained bare noun effect for pooled cognates and noncognates (overall), for noncognates and cognates across Gender Compatibility conditions in the two proficiency groups of the bilingual group. (Bare noun effect = mean RTs of bilingual group minus mean RTs of monolingual group).

For the low-proficient group, statistical analyses revealed that the interaction of Gender Compatibility (only bare nouns) with Experiment Group was not significant in the overall analysis ($F_1(2, 74) = 1.039$, $p = 0.359$; $F_2(2, 107) = 0.625$, $p = 0.537$), marginally significant in the F_1-analysis of the noncognate analysis ($F_1(2, 74) = 2.393$, $p = 0.098$; $F_2(2, 66) = 2.095$, $p = 0.131$), and not significant in the cognate analysis ($F_1(2, 74) = 0.549$, $p = 0.580$; $F_2(2, 38) = 0.354$, $p = 0.704$).

For the high-proficient group, statistical analyses revealed that the interaction of Gender Compatibility (only bare nouns) with Experiment Group was not significant, neither in the overall analysis ($F_1(2, 76) = 0.835$, $p = 0.438$; $F_2(2, 107) = 0.164$, $p = 0.849$), marginally significant in the F_1-analysis of the noncognate analysis ($F_1(2, 76) = 2.673$, $p = 0.076$; $F_2(2, 66) = 0.727$, $p = 0.487$), and not significant in the cognate analysis ($F_1(2, 76) = 0.153$, $p = 0.858$; $F_2(2, 38) = 0.164$, $p = 0.849$).

Summary

There were no significant interaction effects of Gender Compatibility with Experiment Group (except for two marginally significant effects in the F_1-analyses in the noncognate analyses for low- and high-proficient subjects) and thus no evidence for a bare noun gender interference effect.

4.9 Discussion LDT

The results found for the German and Spanish bilingual participants were largely similar. Unlike Lemhöfer, Spalek, and Schriefers (2008), no interaction effects of Gender Compatibility with Prime Type were found in RTs. The tendencies of RTs across Gender Compatibility conditions were also inconsistent so that there was not even a trend into the expected direction. A significant effect of Gender Compatibility for bare nouns was not found, either. Error rates were too low to investigate possible effects. As shown in the monolingual analyses, item-matching seemed to be successful in the case of the Spanish version of the LDT, but less so in the German version.

As explained in the discussion of the PNT (cf. section 4.4), one possible explanation for the failure to find a gender interference effect and the inconsistent RT patterns across the relevant conditions in the bilingual groups is the fact that Prime Type was not a complete within-subjects variable. That way, the RTs for the unprimed items of some subjects had to be compared to the RTs for primed items of other subjects, which probably heavily increased the variance. Furthermore, as also explained in the discussion of the PNT, the exclusion of items due to translation errors can affect the balance of the experimental design despite careful item-matching.

4.10 Discussion Experiment 1

The goal of the present experiment was to investigate gender interference effects in NP processing between a Germanic language with a fairly intransparent gender system and a Romance language with a transparent gender system. An important question to be answered was, whether gender interference can occur between languages from different language families with asymmetric gender systems. Moreover, the role of gender transparency and language proficiency in gender transfer merited further investigation. To this end, a PNT, an offline gender assignment task, and an LDT were conducted. That way, different processing modalities (production vs. comprehension, online vs. offline) could be compared. RTs (only online tasks) and error rates (across all tasks[41]) were measured. Subjects were native speakers of German and Spanish with different proficiency levels in their respective L2 Spanish or German, in addition to monolingual speakers in the two native control groups. Cognates as well as noncognates were used as stimuli. Stimulus conditions included a congruent condition with nouns that were gender-congruent across the two languages, an incongruent condition including nouns that were masculine in one language but feminine in the other, and vice versa, and an incongruent neutral condition including nouns that were neutral in German and masculine or feminine in Spanish. Due to differential findings in the literature for Romance and Germanic languages, interference effects were investigated in NP as well as bare noun

[41] Error rates were only analyzed for the PNT and offline gender assignment task.

processing (in the online tasks). Interference effects for the incongruent conditions were predicted across all three tasks, in RTs (online tasks) as well as error rates. It was hypothesized that transfer effects would be stronger at lower proficiency levels than at higher proficiency levels and that effects would possibly be stronger for cognates than for noncognates.

A gender interference effect was found in the error rates of the PNT for both bilingual groups (only significant by participants). This effect seemed stronger and more reliable for the Spanish bilingual group. In the German bilingual group, this gender interference effect seemed to be somewhat mediated by proficiency. That is, the low-proficient but not the high-proficient group showed an almost significant difference between the congruent and incongruent condition. However, the overall error rate was so low in the German group that this result has to be interpreted with caution. In the PNT as well as in the offline gender assignment task, low-proficient subjects made more errors than high-proficient subjects in both language groups. In the error rates of the offline gender assignment task, a clear transfer effect was found for the Spanish bilingual group, but not for the German bilingual group. Significantly more gender errors were committed in the two incongruent conditions than in the congruent condition. This gender transfer effect in the Spanish bilingual group was not mediated by proficiency. The failure to find a gender transfer effect in the German bilingual group might be due to the German subjects almost performing at ceiling because of the great transparency of the Spanish gender system. For the Spanish bilinguals, on the other hand, problems with gender assignment persisted at all proficiency levels.

In the RTs of the two online experiments, the PNT and LDT, no gender interference effects became apparent, neither for NPs nor for bare nouns. As explained in the discussions of the PNT and LDT, the reason for this might be attributed to a characteristic of the experimental design which decreased the likelihood of observing effects in the RT. In addition, the necessary exclusion and loss of items due to errors further increased the imbalance of the design.

Regarding effects of Cognate Status on gender transfer, no differences between noncognates and cognates were observed in the data where gender transfer effects were found. This is different from findings in the literature where stronger gender interference effects for cognates than for noncognates have been reported for several experimental tasks (Lemhöfer, Schriefers, & Hanique, 2010; Lemhöfer, Spalek, & Schriefers, 2008; Salamoura & Williams, 2007 (only error rates)). However, the lack of finding a difference between cognates and noncognates might be due to the fact that on average, cognates were longer and less frequent than noncognates. The circumstance that Cognate Status could not be included as a factor additionally lowered the likelihood of finding an effect.

In the following, the present results will be further discussed in the light of task demands, differences between L2 German and L2 Spanish, transparency effects, and the present state of the art.

Regarding the importance of different processing modalities (online/offline, production/comprehension; cf. section 2.2), in the present study, gender transfer effects were investigated by means of two online tasks (PNT and LDT) and an offline task (offline gender assignment task). The PNT is a production task and the LDT a comprehension task. L2 gender processing should be more difficult in online tasks than in offline tasks and also more difficult in production than in comprehension studies. In the present study, clear gender transfer effects occurred in the offline gender assignment task of the Spanish subjects. In addition, transfer effects became visible in the error rates of the PNT for the Spanish subjects and to some degree also for the

German subjects. Regarding the online/offline comparison, the emergence of a gender transfer effect for the German subjects in the error rates of the PNT, which was absent in the offline task, might be attributable to the greater task demands in the online tasks. For Spanish subjects, gender interference effects were found in the error rates of the PNT as well as the offline gender assignment task. Thus, no conclusion regarding a higher difficulty in the online task can be drawn. Furthermore, since error rates in the LDT were too low to be analyzed, the comparison between a production and comprehension task cannot be made. In the RTs of the online tasks, no transfer effects became apparent, neither in the PNT nor in the LDT. Therefore, when looking at the RT analyses, no conclusions regarding differences in interference processes between production and comprehension tasks can be drawn, either.

An important finding of the present experiment is that there seems to be a big difference between German and Spanish subjects (even though proficiency matched), or L2 Spanish and L2 German, respectively, regarding L2 gender knowledge and retrieval. In the offline gender assignment task, Spanish subjects made significantly more errors than German subjects. Error rates in article assignment remained high even for high-proficient Spanish subjects (18.8%). Also in the PNT, Spanish subjects made more gender errors than German subjects. Furthermore, the significantly longer RTs in the PNT in German for NP naming compared to bare noun naming suggest that retrieval and selection of a German determiner is problematic for Spanish subjects. Even at higher proficiency levels Spanish subjects continued to have difficulties with determiner selection. Also Lemhöfer, Spalek, and Schriefers (2008) found significantly longer RTs for determiner naming than bare noun naming in their native German speakers of L2 Dutch. This is probably due to the low transparency of the German[42] (present experiment) and Dutch (Lemhöfer, Schriefers, & Hanique, 2010; Lemhöfer et al., 2008) gender systems. The low transparency of these gender systems apparently leaves L2 speakers in doubt at the moment of determiner selection, hampering output. Furthermore, in the offline gender assignment task, effects of frequency were found. This is in line with the findings of Sabourin, Stowe, and de Haan (2006) who found frequency effects in a gender assignment task in L2 Dutch, which also has an intransparent gender system.

For the German subjects in the present experiment, gender assignment and determiner retrieval did not seem to be problematic. In the offline gender assignment task, German subjects were able to perform close to ceiling even at low proficiency levels (4.2 %). In the PNT, the RTs between the bare noun and determiner naming condition did not differ significantly even for low-proficient German subjects. This is probably due to the high transparency of the Spanish gender system which allows L2 speakers of Spanish to quickly retrieve the determiner[43]. The importance of the transparency of the L2 gender system is underlined by the fact that, as stated above, in the offline gender assignment task, a transparency effect for L2 Spanish was found. Significantly less gender assignment errors were committed for transparent nouns than intransparent nouns. Different from the Spanish subjects, word frequency did not affect accuracy rates for the German subjects. Apparently, transparency plays a more important role than frequency in the case of a transparent L2 gender system. The minor

[42] Note that in the present experiment also the fact that German has a three-way gender system might have contributed to the high error rates and general difficulties of the Spanish subjects. As stated before, the odds of making a correct guess are 33 % in German but 50 % in Spanish.

[43] The lack of a difference in RTs between bare noun and determiner naming in L2 Spanish argues against the theory that for transparent L2 gender systems determiners are computed "on the spot", as put forward by Salamoura and Williams (2007). If determiners were computed instead of retrieved, naming times should be longer in the NP condition than in the bare noun condition.

role of frequency is underlined by the fact that accuracy rates for unknown transparent Spanish nouns were still high. Moreover, certainty ratings of the correctness of the assignment were also higher for transparent than intransparent Spanish nouns. Comparable effects of L2 noun ending transparency were found by Bordag and Pechmann (2007) in a PNT and GJT in both RTs and error rates and also by Bordag and Pechmann (2008) in the RTs and error rates of a translation task (cf. section 3.3.2).

These results and my results are in line with findings from the literature showing that processing of irregular nouns is more difficult than processing of regular nouns in L1 (E. Bates, Devescovi, Pizzamiglio, D'Amico, & Hernandez, 1995; Hernandez et al., 2004) and in L2 (Hernandez et al., 2007) because transparent morphophonological cues facilitate processing (Frenck-Mestre et al., 2008; McLaughlin et al., 2010). Intransparent form–function mappings, on the other hand, are difficult for late L2 learners and might even not be acquired at all, leading to fossilization (DeKeyser, 2005). DeKeyser further points out that low reliability and low salience of grammatical features, as it is the case for the German grammatical gender system (cf. section 4.1.1), make mappings difficult to acquire. Also Blom, Polisenska, and Weerman (2008) argue that "morphophonological regularities enable adult learners with extensive immersion in the L2 to perform rather accurately with respect to grammatical gender." (p. 300). This explains the high accuracy of the bilingual German speakers in the offline gender assignment task. However, Blom et al. also state that late learners´ knowledge of gender is purely lexical in nature and that "By using lexicon-driven strategies to learn grammatical gender, late learners can thus show seemingly native behaviour, even though they lack a grammatical representation of grammatical gender." (p. 300). This might lead to problems at greater agreement distances (cf. section 2.2), such as tested in Experiment 2.

In addition, my findings on the importance of transparency for the accuracy in offline gender assignment are in line with the predictions made by the Competition Model on cue strength and cue validity (cf. section 1.5.2). The frequency or availability and reliability of cues are important in (L2) acquisition. As discussed in section 4.1.1, German gender cues are quite unreliable, while gender cues in Spanish transparent nouns are very reliable. Furthermore, the predictions of the Competition Model concerning negative transfer in the incongruent condition due to language competition and positive transfer in the congruent condition were borne out. Regarding the DP Model (cf. section 1.5.1), which postulates the reliance on declarative structures instead of the reliance on procedural structures for late L2 learners, no conclusions can be drawn from the present data.

Furthermore, regarding the gender interference effects observed in the offline gender assignment task, and transfer effects in general, another remark has to be made. Clearly, not all the errors in L2 are made because of gender transfer because then there would be no errors in the congruent condition. So even if there is ample evidence for L1 gender transfer (cf. chapter 3), it apparently cannot account for all of the gender errors and thus the difficulties L2 speakers experience. Consequently, there must also be other reasons for the faulty representations in L2 learners. Regarding this issue, Ortega (2009) remarks that "Thus, many errors that at first blush might be attributed to the influence of the mother tongue can be, in fact, unrelated to the L1 and instead reflect developmental universal processes that have been attested in the acquisition of human language in general (and often in L1 acquisition as well, where no pre-existing knowledge of a specific language can be assumed to influence the process)." (p. 51). Ortega also cites a study by Ellis (1985) who found that the amount of L1 transfer typically accounted for 23 to 36 % of production errors. Hence, it seems that L1 transfer can always only explain a part of the L2 errors committed. Other factors, such as universally similar sequences in the acquisition of certain language features, also play a role. Furthermore, the instability of L2 gender representations in late learners, as observed by Lemhöfer, Spalek, and Schriefers (2008), might be another explanation for the errors occurring in the congruent condition (even though incongruent nouns are more unstable than congruent nouns). In addition, as mentioned in the introduction of chapter 3, Clahsen and Felser (2006a) discuss four main explanatory attempts for L1 and L2 processing differences: lack of relevant grammatical knowledge (i.e., a lack of competence), influence from the L1, cognitive resource limitations, and maturational changes after puberty (p. 564). Accordingly, in the end, a combination of several factors might be at work.

In summary, the present experiment provided evidence for the interaction of the gender systems of a Romance and a Germanic language. This is suggested by the results in the error rates of the PNT and the offline gender assignment task. Furthermore, the transparency effects found in the present experiment raise the interesting possibility that transfer can also be mediated by L2 characteristics. This is especially noteworthy since so far studies have mostly looked at effects of L1 characteristics (Sabourin & Stowe, 2008; Sabourin et al., 2006; Sagarra & Herschensohn, 2011). The influence of L2 characteristics will be further explored in Experiment 2 which investigates whether gender transfer is also possible into the English language. English has a purely semantic and as such an even more transparent gender system than Spanish. The potential importance of L2 transparency in language transfer had already been mentioned by Lemhöfer et al. (2008, pp. 327-328; cf. also section 4.5.3). Nevertheless, transfer from German to Spanish might occur with more intransparent and infrequent Spanish nouns or under circumstances of greater time pressure or greater task demands such as having to compute greater agreement distances (cf. section 2.2). This seems likely because of the fact that German subjects showed gender interference in the studies by Lemhöfer, Schriefers, and Hanique (2010) and Lemhöfer et al. (2008), the transparency effects found for L2 Spanish in the present study. Moreover, another important conclusion is that, as shown in the error rates of the PNT and the offline gender assignment task, gender transfer does not only arise when the L1 and L2 are as similar as German and Dutch in Lemhöfer et al.´s studies but also when the two languages and gender systems are as dissimilar as Spanish and German.

As stated in section 3.3.2, the direct comparison of the results in the offline task with the results in the online tasks should allow for the comparison between the acquisition-based account and the online account put forward by Lemhöfer et al. (2010). The fact that gender interference was observed in the offline gender assignment task of the L1 Spanish subjects shows that these L2 speakers of German had incorrect and possibly unstable gender representations which are biased by L1 influences (acquisition-based account). Even if from this result it cannot be concluded whether the two gender systems are really shared or not, it appears that, stated in terms of the Competition Model, the L2 system is parasitic on the L1 system. So gender transfer seems to at least occur in the acquisition phase (and when the L2 gender system is intransparent), leading to faulty representations in the case of incongruent gender. Possibly because of methodological problems, no further evidence pointing towards a shared bilingual gender system could be obtained in the RTs of the PNT and LDT.

Overall, the fact that evidence for gender interference was provided in an offline task suggests that, as explained by Lemhöfer, Spalek, and Schriefers (2008), the high error rate in L2 gender tasks might in some cases actually stem from incorrect gender representations (acquisition-based account) rather than from online interference effects (online account) even though "this imperfect acquisition is heavily biased by L1" (Lemhöfer et al., 2008, p. 328). Hence, even if the RTs revealed no online gender interference effects, the results obtained in the error rates of the PNT and offline gender assignment task suggest that the two language systems of a bilingual interact at least in the acquisition phase. This means that gender transfer between a Germanic and a Romance language is possible. Furthermore, regarding the influence of transparency on gender transfer it can only be concluded that transparency at least helps with gender acquisition.

What do the present results mean in the light of the literature discussed in section 3.3.2? First of all, according to Costa, Kovacic, Franck, and Caramazza (2003), regarding the bilingual gender system, two hypotheses can be put forward: the “gender-integrated” (p. 182) and the “language-autonomy” (p. 183) view. As stated in the previous paragraph, even though no online interference effects could be obtained in the error rates, the present results point towards a shared gender system rather than a separate language system. Similar to Bordag and Pechmann (2007), who found gender congruency effects in a PNT between L1 Czech and L2 German, I also found gender congruency effects between two different language families. However, contrary to the language pairs used in the present experiment, Czech and German both have three-way and thus symmetric gender systems. This is also the case for L1 Greek and L2 German used in the translation study by Salamoura and Williams (2007). So the present study shows that gender transfer is not only possible across language families but also in the case of asymmetric gender systems. Furthermore, the present study confirms the important role of noun ending transparency in mediating transfer effects found by Bordag and Pechmann. Lemhöfer, Schriefers, and Hanique (2010) and Lemhöfer et al. (2008) had found transfer effects between L1 German and L2 Dutch, in the RTs and error rates of an LDT and PNT as well as in the error rates of an offline gender assignment task. German and Dutch belong to the same language family but have asymmetric gender systems. The present experiment replicated the gender transfer effect in the error rates of the PNT and offline gender assignment task (RTs of German PNT) between two languages from different language families with asymmetric gender systems. Contrary to Paolieri et al. (2010), who had observed gender transfer effects in a PNT in bare nouns as well as NPs, no online interference effects could be observed for neither of the two structures. Therefore, no conclusions regarding the lexical vs. syntactic hypothesis (section 3.3.2 Studies looking at gender congruency effects; Paolieri, Cubelli, et al.; 2010, p. 2) can be drawn. Nevertheless, it seems a worthwhile approach to investigate interference effects in NPs as well as bare nouns when investigating the interaction of the gender system of a Germanic and Romance language.

In conclusion, gender transfer between languages from different language families, a Romance and German language, with asymmetric gender systems is possible. Regarding the direction of language transfer, it seems to be possible in both directions. Because of the great difference in performance between the German and Spanish subjects especially in the offline gender assignment task, my data further suggest that this L1 transfer is mediated by L2 factors: L1 transfer is smaller if the L2 gender system is transparent and greater if the L2 gender system is intransparent. Possibly because of these strong transparency effects, there was only little evidence for proficiency effects in language transfer. Effects of Cognate Status could not be determined. Due to methodological problems, no differences between language production and comprehension could be observed. But there was some evidence towards a greater difficulty in an online task compared to an offline task.

As stated above, it is possible that L1 gender transfer effects into an L2 with a transparent gender system, such as Spanish, might become visible in online tasks taxing working memory resources, under greater time pressure and in the context of greater agreement distance. In Experiment 2 I will look at L1 gender interference effects for low-proficient German subjects into an L2 with an even more simple gender system than Spanish, namely, English, during sentence processing and at a greater agreement distance. The present data suggest that there might be a trade-off in gender transfer between task difficulty and L2 gender system transparency. In Experiment 2, this possibility will be further explored, including proficiency effects.

5. Experiment 2

5.1 Introduction

Unlike Experiment 1, which investigated L1 gender transfer effects in NP processing, Experiment 2 investigates L1 gender transfer effects in an L2 sentence processing context. This time ERPs are measured, as ERPs can give information on the time course of syntactic and semantic information integration during online processing.

In addition, the aim of this experiment is to answer the question whether L1 gender transfer is also possible into a language without a grammatical gender system, namely, English. In terms of complexity of the gender systems of the languages tested in the present thesis, one could say that German has the most "difficult" gender system, that is, the least transparent one. Often, noun endings do not provide a cue to the noun´s gender and the rules regarding noun terminations are complex (cf. section 4.1.1). Moreover, German has a three-way gender system. The Spanish gender system, on the other hand, is fairly transparent and simple, and it is also only a two-way gender-system (cf. section 4.1.1). The English language has the simplest gender system of the languages tested in the present thesis. English is a Germanic language with Romance influences. The English gender system lacks syntactic gender and is purely based on semantic criteria (Corbett, 1991, p. 12). Human beings can have masculine or feminine gender, for the rest, all nouns, including animals, are neutral. Due to this overlap with biological sex, it is very easy to infer the gender of a noun. There are a few exceptions, such as ships being thought of as feminine and pets being referred to by their sex. Furthermore, English has a pronominal gender system which means that gender surfaces only in personal, possessive and reflexive pronouns. The definite and indefinite articles, as well as demonstratives and adjectives take just one form for all genders. In brief, the English gender system is so simple that it can be basically summarized in one rule, namely, that gender overlaps with sex. Therefore, in terms of transparency, the English gender system could also be considered an "extremely transparent" gender system. As we have seen in Experiment 1, transparency of the L2 gender system played a role in the correctness of gender assignments and also for L1 gender transfer. L1 gender transfer was stronger for L2 German than for L2 Spanish. Furthermore, for L2 Spanish, tendencies towards transparency effects were observed. I now investigate whether gender transfer is also possible into an L2 without a gender system, or if L1 gender transfer effects will be completely absent due to the high transparency of the gender system.

Compared to the bulk of research on language transfer, two things are special about this experiment: First, the fact that it investigates the influence of L2 characteristics, that is, whether gender transfer is possible into an L2 with a very simple or transparent gender system. Especially in combination with the first experiment, this gives insights into the role of L2 transparency. It is also important to note that, regarding gender transfer, so far studies have mostly looked at effects of L1 characteristics (e.g., Lemhöfer, Spalek, & Schriefers, 2008; Sabourin, Stowe, & de Haan, 2006; Sabourin & Stowe, 2008; Sagarra & Herschensohn, 2011). As stated by Frenck-Mestre, Foucart, Carrasco, and Herschensohn (2009) "Over and above the possible influence of native language parameters on the L2 acquisition and processing of gender concord, which has indeed been a major focus of SLA research […], it is of interest to examine factors within the L2 that may play a role." (p. 88). Nevertheless, while previously the focus was very much on L1 factors affecting transfer (cf. chapter 3), by now a few studies have started to investigate the role of L2 factors in transfer. Examples are the studies discussed earlier which found effects of L2 gender transparency cues in gender transfer (Bordag et al., 2006; Bordag & Pechmann, 2007, experiment 3 and 4; Frenck-Mestre et al., 2009, experiment 2).

In addition to the influence of L2 factors, a second point of interest in language transfer is whether transfer of an L1 feature is possible if this feature does not exist in L2. Furthermore, while in the previous experiment also offline gender transfer effects due to erroneous representations were investigated (cf. section 4.5), in the present experiment, it was assumed that all participants were familiar with the basic rule that English lacks syntactic gender. In other words, participants´ L2 gender representations for the inanimate objects used in the present experiment should be correct (inanimate objects = neutral gender). Thus, importantly, if in the present experiment transfer effects are found, they support the online account of gender transfer rather than the representational account.

Few studies so far have investigated whether transfer of a morphosyntactic feature that is present in L1, but not in L2, might affect processing in an L2. Barto-Sisamout, Nicol, Witzel, & Witzel (2009) conducted a self-paced reading task with Spanish learners of English, a comparison group of Chinese learners of English (because Chinese lacks the morphosyntactic markings in question), and a native control group. Their aim was to test the influence of the similarity and presence/absence of morphosyntactic rules in L1 (Spanish) and L2 (English). One condition was the "similar but different" condition, in which the L1 and L2 both encompass the same morphosyntactic concept but with different rules, namely, possessive pronouns. They took advantage of the fact that in English, possessive pronouns agree in person and number with their antecedent, while in Spanish, possessive pronouns agree in person with their antecedent and in number with the noun they modify. For example, in the sentence *The author$_{\text{-sing}}$ wrote her$_{\text{-sing}}$ articles$_{\text{-plu}}$ at the coffee shop./ La autora$_{\text{-sing}}$ escribió sus$_{\text{-plu}}$ artículos$_{\text{-plu}}$ en la cafetería.* the English pronoun *her* is singular because its antecedent *author* is singular even if the modified noun *articles* is plural. The Spanish pronoun *sus,* on the other hand, has to be plural, because it has to agree in number with the modified noun *artículos*, which is plural. The second condition, the "L1+L2-" condition, tested the influence of having a rule in L1 that does not exist in L2, namely, personal and non-personal direct objects. In Spanish, personal direct objects require the preposition *a*: *El artista vio* ***a*** *la bailarina en la discoteca de moda* (The artist saw the dancer at the trendy nightclub). But non-personal direct objects do not require a preposition: *El artista vio el cuadro en la discoteca de moda.* (The artist saw the painting at the trendy nightclub.). It was predicted that when reading in their L2 English, native Spanish speakers would expect some kind of marking for personal direct objects, too, resulting in slower reading times. Only a trend towards slower reading times was found in the "L1+L2-" condition, for the subgroup of Spanish - English bilinguals who started learning after age 10. Therefore, the authors noted that the lack of finding clear interference effects in the two conditions of interest might also be due to the fact that, overall, their bilingual subjects had a very high proficiency. This experiment shows that it might be difficult to obtain transfer effects for a morphosyntactic feature that is present in L1 but absent in L2, especially for early bilinguals and at high proficiency levels. At a later point, I will describe the characteristics of my experimental design that are thought to increase the likelihood for observing transfer effects in the present experiment. Next, I will explain the characteristics of anaphor resolution and more specifically, pronoun resolution, which was investigated in the present experiment.

The main reason for investigating pronoun resolution was that in the chosen L2, English, gender only surfaces in pronouns. Furthermore, it was intended to increase the agreement distance between the noun and its agreeing element in comparison to the first experiment, in order to investigate effects of agreement distance in gender transfer. But what does anaphor and pronoun resolution actually entail from a processing perspective? A definition is provided by Huang (2000): “The term ‘anaphora’ is derived from the Greek word ἀναφορά which may mean ‘carrying back’. In contemporary linguistics, it is commonly used to refer to a relation between two linguistic elements, wherein the interpretation of one (called an anaphor) is in some way determined by the interpretation of the other (called an antecedent) […]. Linguistic elements that can be employed as an anaphor include gaps (or empty categories), pronouns, reflexives, names and descriptions.” (p. 1). The ability of anaphora resolution, that is, the ability to select the correct referent from a list of possible candidates, is essential for understanding coherent discourse or longer passages of text.

For the present experiment, pronouns were chosen as anaphors. Lamers, Jansma, Hammer, and Münte (2008) point out that: “Personal pronouns play an important role in discourse understanding. They form cohesive links between sentences and sentence fragments by referring back to a linguistic element, the antecedent, in a so-called co-referential relationship.” (p. 2). This is illustrated by the following sentence pair: *The girl plucked an apple from the tree and gave <u>it</u> to the boy. While <u>he</u> was eating <u>it</u>, <u>she</u> plucked <u>another one</u> for <u>herself</u>.* In order to understand this simple story, several anaphora (underlined) have to be resolved correctly and referred to the correct antecedent: *it* referring to *the apple*, *he* referring to *the boy*, *she* referring to *the girl* and so on. In order to be able to do so, all the antecedents have to be retained in short-term memory in order to be retrieved again upon reading the anaphor. In gendered languages, gender has an important function in pronoun resolution. Lamers et al. further state that “To establish a cohesive link between the pronoun and its antecedent (i.e. the man, the woman, the child) the pronoun inherits the gender (masculine, feminine, or neuter) and number (singular and plural) characteristics of the antecedent” (p. 2). So, for speakers of languages with grammatical gender, gender serves as an important cue in pronoun resolution. In English, a language lacking syntactic gender, the personal pronouns *he* and *she* can (almost) only refer to animate entities. In German, on the other hand, the pronouns *he* and *she* can also refer to inanimate things. As discussed in chapter 3 and as predicted by the Competition Model (cf. section 1.5.2), L1 cues can affect L2 processing. This poses the possibility of investigating whether L1 speakers of German would suffer from interference of the German syntactic gender value of an object when resolving pronouns in English. For this reason, pronoun resolution seems to be a suitable syntactic structure to study gender transfer effects from German to English.

However, as indicated by the results of the first experiment, gender transfer seems to be less likely when the L2 has a very transparent gender system. Taken together with the result of Barto-Sisamout et al. (2009) this suggests that, in the second experiment, it could be difficult to make potential transfer effects visible. Nevertheless, several characteristics of the present experiment should increase the likelihood of observing transfer effects: First of all, transfer effects are observed in a sentence context and the structural distance of the agreeing elements was greater than in the first experiment (cf. section 2.2 on effects of agreement distance). Second, since transfer effects seem to be greater at low proficiency levels (cf. chapter 3), very low-proficient subjects were recruited for the experiment. Third, a measure that is more sensitive than RTs and errors rates was employed by using ERPs. ERPs can reveal effects that do not become visible in behavioral measures. In addition, ERPs give detailed information on the time course of language processing (e.g., McLaughlin, Tanner, Frenck-Mestre, Valentine, & Osterhout, 2010). In the following, other studies investigating L2 influences in transfer and pronoun processing will be discussed, as well as studies investigating gender transfer or congruency effects with ERPs. The advantages of the ERP method will also be described briefly.

Some studies have shown that L1 influences can have an impact on L2 pronoun processing. For example, Antón-Méndez (2010) conducted a semi-spontaneous production task and found that Spanish speakers made more pronoun errors, that is, *he*/*she* confusions in English than French speakers. Gender errors were also more common than number errors in the pronoun production of Spanish speakers. The authors reasoned that this was due to the fact that Spanish is a pro-drop language where pronouns can oftentimes be omitted unless for the purpose of making an emphasis so that the gender information does not have to be retrieved. Number information, in contrast, has to be made available for verb inflection. These results show that the L1 also exerts an influence in pronoun production and can lead to gender errors. This study is in line with other studies showing that morphosyntactic characteristics of the L1 influence morphosyntactic processing in L2. It furthermore shows that this influence also takes place in pronoun processing involving grammatical gender errors.

So far, studies have mostly investigated the influence of L1 (morpho)syntactic characteristics on L2 processing, as well as the influence of the presence or absence of certain features in L2 (cf. section 3.2.2). An example is the study by Antón-Méndez, described above. Frequently, the conclusion has been drawn that L2 learners transfer L1 processing strategies and that L2 learners have problems with L2 morphosyntactic features that are not present in their L1. However, the question what happens when a feature is present in L1 but not in L2, and has to be “switched off” or suppressed when processing L2, has not been studied much. An exception constitutes the study by Conklin, Dijkstra, and van Heuven (2007). These authors obtained L1 gender transfer effects from Dutch to English also in a fairly natural (auditory) sentence processing task, and they also used a task involving pronoun resolution, as in the present experiment. A visual world eye-tracking paradigm was used. Conklin et al. displayed a scene on a screen which contained a cartoon character, which has biological gender in both English and Dutch, and an inanimate object which has grammatical gender (common or neuter) in Dutch and neuter gender in English. The cartoon character and the object either agreed in gender in Dutch or not. Sentence pairs like *The tractor will be driven by Donald Duck. He is in the other field*. were used. For native Dutch speakers, the pronoun *he* could be ambiguous because in their L1 it would be unclear whether it refers to the tractor (which is masculine in Dutch) or to Donald Duck (who is also masculine). If, however, a feminine cartoon character, like Daisy, was displayed together with the tractor, there would be no ambiguity. The results showed that Dutch subjects exhibited significantly more looks to the object than cartoon characters in ambiguous situations, that is, when the object and cartoon character agreed in gender. In addition, the object was looked at significantly more often in ambiguous than in unambiguous situations. No differences between conditions were observed for a monolingual English control group. Thus, as revealed in this eye-tracking experiment involving auditory sentence processing, L1 gender information is activated and can interfere in the process of L2 pronoun resolution, even if the L2 lacks gender.

The online measurement of ERPs was chosen as experimental method for the present experiment. The advantage of measuring ERPs is that one obtains a continuous measurement of language processing unfolding in the brain. The measurement of RTs or error rates gives mainly information about the result of processing and little information on the process itself. From these behavioral measurements, the cognitive processes have to be inferred, such as for example, longer RTs pointing to processing difficulties. With ERPs, on the other hand, one can investigate at what time point a processing difficulty occurs. Additionally, the established components allow the researcher to gain more insights on the nature of the difficulties.

A further advantage of the ERP technique is the fact that ERPs often can be collected in addition to behavioral measures such as RTs and error rates. Since ERPs are a more sensitive measure, they can reveal processing difficulties (e.g., in L2 speakers) even if performance in terms of behavioral measures appears to be at ceiling. ERPs have previously been used successfully to investigate gender transfer and gender congruency effects. The importance of investigating gender transfer effects with electrophysiological measures has been pointed out by Kotz (2009, p. 69): "While the transfer phenomenon has been extensively discussed in the psycholinguistic behavioural L2 literature, to date it has barely been considered in neurophysiological and neuroimaging investigations [...]. Such a discrepancy clearly indicates the need to bring psycholinguistic theory and neural investigations of L2 syntactic acquisition closer together." In the following, I will review some of the studies investigating gender transfer effects with ERPs in order to show how what the present experiment can contribute to the state of the art.

In the earlier cited studies by Foucart and Frenck-Mestre (2011) and Frenck-Mestre et al. (2009) (cf. sections 2.1 and 3.3.1, respectively), it was observed that native speakers of German were insensitive to gender violations in their L2 French when these violations involved plural forms. This is probably due to the fact that plural forms in German do not require gender agreement. Furthermore, Foucart and Frenck-Mestre also observed proficiency effects for determiner–noun agreement violations as some L2 subjects showed more sensitivity to the violation than others. In addition, for a subgroup of the L2 participants a gender congruency effect was observed, that is, differences in P600 amplitude depending on whether nouns had the same or a different gender in French and German. An attenuated P600 for agreement violations involving French nouns that had a different gender in German was observed. These results suggest that ERPs are sensitive to L1 influences in L2 gender processing and gender congruency effects.

L1 gender transfer effects in ERPs were also shown by Ganushchak, Verdonschot, and Schiller (2011). However, quite a different method and task than in the previous ERP experiments was employed. What is also special is that, just as in the present experiment, they investigated L1 gender transfer into a language that lacks gender, namely English. Subjects were native speakers of Dutch who were highly proficient in their L2 English. Their task consisted in making gender classifications to Dutch words (printed in white) and color decisions to English and Dutch words (printed either in green or blue) by means of pressing a button. For Dutch common gender, the left key had to be pressed, while for Dutch neuter gender the right key had to be pressed. For color decisions (green/blue), response buttons were counterbalanced among subjects. This setup resulted in congruent trials, when the color as well as the gender decision required the same response (e.g., left button for a green word and the green word was neuter) or incongruent trials, when the color and the gender decision did not require the same response (e.g., right button for a blue word and the blue word was neuter). For analyses, only trials including color decisions were relevant. No effects of congruency were found in the RTs, but effects were found in the error rates: More errors were committed in situations of response conflict (i.e., incongruent trials) than when the trial was congruent, for Dutch words as well as for English words. Regarding ERP analyses, the relevant component was the event-related negativity (ERN), which is observed in situations of response conflict and error detection (p. 106). The results showed that for both languages the ERN was greater in incongruent trials than in congruent trials. Hence, the "gender-color conflict" was not only observed for Dutch, but also for English, which lacks grammatical gender. Since the response conflict in the English trials can only stem from the gender conflict in Dutch, this shows that the L1 gender information from Dutch is transferred into the L2 English. However, it has to be noted that the task of the present experiment was not very natural and the stimuli used had little in common with everyday language processing. As also remarked by the authors, the mixing of trials might have led to the strategy of retrieving color and gender information in all trials. Furthermore, the strategy of retrieving syntactic gender in all trials might have been reinforced by the fact that because of the explicit gender decisions that had to be made, the focus of the experiment was very much on gender. Thus, the second experiment of my thesis has the advantage of investigating gender interference into the L2 English in a more natural processing context, namely, sentence processing. Moreover, components such as the N400, LAN, and P600 observed in sentence processing tasks can give more information on the time course and nature of processing, native-likeness of processing, and the sources of potential difficulties, than the ERN. Besides, as also noted by the authors, the mixing of languages in the experiment greatly abets language transfer, leaving the question open whether subjects would also experience interference when they are not in a language-mixing mode. Ganushchak et al. point out that "Future research will need to determine whether gender transfer between languages can also be observed in a nonmixed-language context even when one language has no gender system." (p. 110).

An example of an ERP study investigating gender transfer effects in more natural sentence processing is the experiment conducted by Midgley, Wicha, Holcomb, and Grainger (2007). They tested native French speakers in their L2 English in a sentence comprehension task. Comprehension questions had to be answered in 20 % of the trials. The authors took advantage of the fact that in French, possessive pronouns are gender-marked and have to agree with the noun they modify, that is, the item that is owned. The possessive pronouns do not have to agree with the gender of the referent, that is, the person the item belongs to. That way, the sentence *Paul dropped his apple on the floor* might be perceived as incorrect or incongruent by native French speakers because *his* is masculine but *apple* is feminine in French. The sentence *Mary dropped her apple on the floor.*, on the other hand, would probably be perceived as perfectly congruent and correct. Indeed, a processing difference between these congruent and incongruent sentences was observed in the ERPs. Compared to "congruent sentences", "gender incongruent" sentences elicited a positivity between 400 and 700 ms. Yet, this positivity did not completely resemble the classical P600 component, as it was most pronounced at anterior right sites[1]. So Midgley et al. (2007) were able to show that gender congruency between L1 and L2 resulted in some kind of processing difference, however, this difference was not manifested through an established component. Nevertheless, this experiment suggests that the manipulation conducted in the present experiment might be successful in eliciting an effect at the level of ERP components.

In summary, there might be influences of L1 morphosyntactic features on L2 processing even when the L2 lacks this feature (Barto-Sisamout et al., 2009), there is an influence of morphosyntactic similarity on errors in pronoun production (Antón-Méndez, 2010), there are gender transfer effects from a gendered L1 into an ungendered L2 in an auditory sentence processing task using eye-tracking (Conklin et al., 2007), and also ERP results showing influences of L1 grammatical gender in L2 sentence processing (Foucart & Frenck-Mestre, 2011; Frenck-Mestre et al., 2009). Furthermore, there are also two ERP studies that have observed influences of L1 gender transfer even when the L2 lacks gender (Ganushchak, Verdonschot, & Schiller, 2011; Midgley et al., 2007). However, these two ERP studies employed tasks that were quite unnatural (Ganushchak et al., 2011), failed to observe the expected typical P600 (Midgley et al., 2007), or the ERP component in question (the ERN) gave

[1] Note that, as Kaan and Swaab (2003) suggest, a frontally distributed P600 might indicate the revision of non-preferred ambiguous structures as opposed to the posterior distributed P600 which indicates repair processes.

little information on the nature of the processing difficulties (Ganushchak et al., 2011). The present experiment complements the weak points and questions left open by these studies. In the following, the predictions regarding the ERP and behavioral measures will be described.

5.1.1 Predictions

As discussed in section 1.3.2, the most important component found in response to L2 (morpho)syntactic violations is the P600. The P600 is reliably observed in native speakers, and sometimes also in L2 speakers, in response to syntactic violations (Friederici et al., 2002; Guo et al., 2009; Hahne & Friederici, 2001; Hahne, 2001; Mueller et al., 2005; Pakulak & Neville, 2011; Rossi et al., 2006) and morphosyntactic violations (Chen et al., 2007; Hahne et al., 2006; Rossi et al., 2006). It has furthermore been shown to be sensitive to gender agreement violations (Frenck-Mestre et al., 2009; Sabourin & Stowe, 2008) and pronoun violations (Hammer, Jansma, Lamers, & Münte, 2005; Schmitt, Lamers, & Münte, 2002). In some cases, in higher-proficient bilinguals, a LAN is also found in response to morphosyntactic violations (cf. section 1.3.2). However, for the present experiment, my predictions mainly concern the P600, because only low-proficient subjects are tested in the present experiment.

Furthermore, as stated in sections 1.3.2 and 1.4.2, sometimes late bilinguals in early stages of L2 learning show an N400 instead of a P600 (Guo et al., 2009; McLaughlin et al., 2010; Morgan-Short, Steinhauer, et al., 2012; Osterhout et al., 2006; Steinhauer et al., 2009), indicating more semantic and shallow processing strategies instead of profound grammatical parsing as observed in natives and near-natives by way of a P600 component. Therefore, a potential N400 was also be investigated if visual inspection gave reason for this.

Before describing the more specific predictions made for the present experiment, I will give an overview of the stimulus sentences for the different conditions. Stimuli were sentences of the following type: *This is a bus. It/*he/*she is slow and crowded.* or *This is a banana. It/*he/*she is yellow and sweet.* (cf. also Table 5.1). Importantly, the antecedent of the pronoun in the second sentence was the (inanimate) noun in the introductory sentence. The German translation equivalent of the antecedent had either masculine (e.g., *bus*) or feminine (e.g., *banana*) grammatical gender.

That way, three “pronoun conditions” were obtained:

- In the “correct condition”, the baseline condition, pronouns were correct in English but would be incorrect for the German translation (*it* – *“es”*).
- In the “pseudocongruent condition”, the critical condition, pronouns were incorrect in English but correct for the German translation (*bus* and *he*/*banana* and *she*).
- In the “incongruent condition”, pronouns were incorrect for both languages (*bus* and *she*/*banana* and *he*).

<table>
<tr><td colspan="2" rowspan="2"></td><td colspan="2">Grammatical gender of German translation equivalent in the introductory sentence:</td></tr>
<tr><td>a) masculine:
This is a bus.</td><td>b) feminine:
This is a banana.</td></tr>
<tr><td rowspan="3">Conditions</td><td>Correct condition</td><td>It is slow and crowded.</td><td>It is yellow and sweet.</td></tr>
<tr><td>Pseudocongruent condition = critical condition</td><td>*He is slow and crowded.</td><td>*She is yellow and sweet.</td></tr>
<tr><td>Incongruent condition</td><td>*She is slow and crowded.</td><td>*He is yellow and sweet.</td></tr>
</table>

Table 5.1 Examples for stimulus sentences across Pronoun conditions with a) masculine and b) feminine German translation equivalents.

For the behavioral measures, the following results were predicted. Because L1 gender transfer was expected in the pseudocongruent condition, the highest error rates and slowest RTs[2] were expected in this condition. In the pseudocongruent condition, L1 and L2 syntactic information contradict each other, that is, the sentence would be correct in L1 but incorrect in L2. The lowest error rates and fastest RTs, on the other hand, were expected in the incongruent condition. In this condition, L1 and L2 syntactic cues “agree” in that the sentence is incorrect. According to the Competition Model, this is a case of “cue summation” (Hernandez, Li, & MacWhinney, 2005, p. 4). Therefore, no competition between L1 and L2 arises.

Regarding the ERP results, for the two incorrect conditions, a P600 was expected to occur after the violation, that is, the pronoun violation, in comparison to the correct (baseline) condition. The most pronounced P600 was expected for the incongruent condition and the smallest P600 for the correct condition. If gender transfer occurs, the pseudocongruent condition should be processed significantly differently from the incongruent condition. The amplitude of the P600 should be intermediate between the incongruent and congruent condition. On the other hand, if there is no gender transfer effect, an equally pronounced P600 should be observed for the incongruent and pseudocongruent condition with both differing from the congruent condition. This pattern would be expected for a native control group free of influences from German syntactic gender.

[2] Note that for reasons explained in section 5.2.4, after visual data inspection I decided not to conduct the RT analysis.

Regarding the correct condition as the baseline condition, it has to be noted that due to the fact that this condition would nevertheless be incorrect if translated literally to German, it might not be completely free of L1 gender transfer. Thus, it is still possible that a small P600 effect is observed, which would be different if the correct condition was correct in both languages. However, this effect should be rather small. Moreover, since the critical question is whether the pseudocongruent condition will be processed differently from the incongruent condition, the central prediction should not be affected by this circumstance.

5.2 Method

5.2.1 Participants

30 subjects with German as L1 and with low English proficiency (A1 - B1[3]) participated in the experiment. The data of two participants were excluded from the analysis as one of them had misunderstood the instructions and the other subject´s electrode impedances were higher than 5 kΩ. 28 subjects entered the data analysis (8 male, 20 female). Subjects were recruited via the subject database of the Humboldt University of Berlin and were reimbursed for participation. They received € 20 for participating in the EEG session and € 5 for participating in the screening session. They were paid after successful completion of both sessions.[4] Regarding level of education, 21 were students currently enrolled at a university or had studied at a university before. 24 subjects had a diploma from a German secondary school qualifying for university admission ("Abitur"), three had a General Certificate of Secondary Education (certificate usually taken after the fifth year of secondary school; "Realschulabschluss") and one subject did not specify. Age ranged from 20 to 34 years (mean = 26.14 years). All participants were right-handed and indicated (self-report) that they carried out 70 - 100 % (mean 88 %) of activities with their right hand. All subjects had normal vision or corrected to normal vision and no neurological or psychiatric impairment. They were native German speakers and spoke English at beginner´s level. As late L2 learners they had started learning English in secondary school at age 10 or later. The only exception was one subject who claimed to have had some minimal contact with the English language at age 4 already. For 22 subjects, English was the first foreign language they had learned; for 6 subjects it was the second foreign language. None of the subjects was proficient in another foreign language except for one subject who stated that she was also quite proficient in French, though less proficient than in English. All participants filled in a language history questionnaire that included questions about their use and level of English and other foreign languages. An overview of the data is given in Table 5.2.

[3] Definitions of the lower proficiency levels A1-B1 according to the CEFR (Council of Europe, 2001, p. 264):
A1: "Can understand and use familiar everyday expressions and very basic phrases aimed at the satisfaction of needs of a concrete type. Can introduce him/herself and others and can ask and answer questions about personal details such as where he/she lives, people he/she knows and things he/she has. Can interact in a simple way provided the other person talks slowly and clearly and is prepared to help."
A2: "Can understand sentences and frequently used expressions related to areas of most immediate relevance (e.g. very basic personal and family information, shopping, local geography, employment). Can communicate in simple and routine tasks requiring a simple and direct exchange of information on familiar and routine matters. Can describe in simple terms aspects of his/her background, immediate environment and matters in areas of immediate need."
B1: "Can understand the main points of clear standard input on familiar matters regularly encountered in work, school, leisure, etc. Can deal with most situations likely to arise whilst travelling in an area where the language is spoken. Can produce simple connected text on topics which are familiar or of personal interest. Can describe experiences and events, dreams, hopes and ambitions and briefly give reasons and explanations for opinions and plans."
[4] Participants who did not fulfill the requirements received € 5 after the screening session.

	Mean	SD	Min	Max
Age	26.14	4.26	20	34
Age of acquisition of English	10.84	1.58	4	13
Months spent in English-speaking country	0.43	2.27	0	12
Number of foreign languages spoken	1.86	0.76	1	3
Reading frequency in English*	2.57	1.81	0	5
Speaking frequency in English*	2.61	1.55	1	5
Frequency of watching English TV or listening to English radio*	2.64	2.68	0	10
Frequency of writing in English*	1.79	1.79	0	5
English reading proficiency†	4.21	1.10	2	6
English speaking proficiency†	3.93	1.39	1	5
English writing proficiency†	3.57	1.57	1	7

* Frequency measures were indicated on a scale ranging from 0 (not at all) to 10 (very often).
† Proficiency measures were indicated on a scale ranging from 0 (very bad) to 10 (very good).

Table 5.2 Overview of participants´ metadata as collected in the language history questionnaire.

5.2.2 Material

48 experimental sentence pairs in English, such as *This is a bus. It/*he/*she is slow and crowded.* or *This is a banana. It/*he/*she is yellow and sweet.* were created. A complete list of stimulus sentences can be found in section 10.2. Sentences were based on inanimate concrete nouns. Only cognate nouns, that is, nouns with the same meaning which are orthographically and/or phonologically very similar in the two languages were used[5]. The German translation equivalent of the subject in the introducing sentence would have either masculine (e.g., *bus*) or feminine (e.g., *banana*) grammatical gender. 24 experimental sentence pairs included nouns that were masculine in German, 24 included nouns that were feminine in German. The pronoun in the second sentence referred to the noun in the introducing sentence. Each sentence had one grammatically correct form, with the anaphor *it* referring to the subject of the previous sentence and two ungrammatical forms, with the anaphor *he* or *she* referring back to the subject of the previous sentence. In addition, simple adjectives, that is, adjectives that were thought to belong to the basic vocabulary of an L2 learner, were used in the second sentence.

144 filler sentences involving a different type of grammatical violation, a violation of gender stereotypes or no violation were included for different reasons:

[5] This was done in order to increase the likelihood of observing transfer effects, as cross-language effects have been shown to be stronger for cognates than noncognates (cf. section 3.3.2).

- 24 fillers with other grammatical violations than in the experimental sentences, namely grammatical number violations, such as *This is a computer. *He/*she/it *are dusty and broken*. were included in order to prevent subjects from becoming overly sensitized to the gender violations.
- 72 filler sentences involving people (48 correct, 24 incorrect) by denoting professions[6], such as *This is a beautician. *He/*she/it is hungry and thirsty.* were included, also with the aim of disguising the real experimental question and in order for the pronouns *he* and *she* to be correct in some cases. The same adjective pairs were rotated across sentences so that the same adjective pairs could appear with male as well as female stereotyped nouns.
- 48 correct sentences, such as *There are two elephants. They are walking in the woods.* were included to make up for the imbalance of correct and incorrect sentences in the experiment. Because of the structure of the experiment and the nature of the English and German gender systems, otherwise there would have been many more incorrect than correct sentences (cf. Table 5.3).

An overview of the experimental and filler sentences can be found in Table 5.3. Note that over the whole experiment, the number of masculine and feminine German translation equivalents was balanced.

Type		Example	Gender of German nouns	Pronoun structure	Grammatical correctness
Experimental sentences (48 correct, 96 incorrect)		*This is a bus. *He/*She/It is slow and crowded.*	24 masculine, 24 feminine, no neutral nouns	48 x *It is* 48 x *He is* 48 x *She is*	48 correct 96 incorrect
Filler sentences (96 correct, 48 incorrect)	a) Grammatical violations fillers	*This is a computer. *He/*she/it *are dusty and broken.*	2 masculine, 2 feminine, 4 neutral	8 x *It are* 8 x *He are* 8 x *She are*	24 incorrect
	b) People fillers	*This is a cook. He/she/*it is nice and friendly*.	36 masculine, 36 feminine	24 x *He is* 24 x *She is* 24 x *It is*	48 correct 24 incorrect
	c) Correct fillers	*There are two elephants. They are walking in the woods.*	16 masculine, 16 feminine, 16 neutral	48 x *They*	48 correct
Total			130 masculine, 130 female, 28 neutral	80 x *He* 80 x *She* 80 x *It* 48 x *They*	144 correct, 144 incorrect

Table 5.3 Examples for sentences and overview of the distribution of correct and incorrect sentences, pronoun structure and German gender of the translation equivalent across experimental and filler sentences.

[6] Note that originally, it was planned to investigate the effects of stereotypes in L2 processing with these filler sentences. Therefore, strongly stereotypically female or male professions such as, for example, hairdresser or engineer were chosen.

Experimental sentences were evaluated by three native English speakers who judged whether the sentences sounded intuitively right (except for the grammatical errors). Additionally, 8 practice sentences were created, approximately reflecting the distribution of correct and incorrect sentences and German gender across experimental and filler sentences. Overall, as in Experiment 1, the aim was to select easy words that would also be known by beginning L2 learners. As indicated by a translation task and familiarity ratings (carried out after the experiment, cf. section 5.2.2), the selected items were quite well-known by the participants. Items were translated correctly on average 92 % of the times. All the participants translated at least 79% of the items correctly. Except for four items, all the items were translated correctly by at least 80% of the participants. The mean familiarity rating was 8.8 (out of 10) with a standard deviation of 2.53. The lowest mean familiarity score per participant was 7.5. Only five items had a familiarity score smaller than 7. The mean familiarity rating was 8.8 with a standard deviation of 2.53. 92% of the items were translated correctly. There was a significant relationship between familiarity ratings and correctness of translation ($X^2 = 609.485$, df = 10, $p < 0.001$).

Experimental sentences were pseudorandomized with the help of the program Mix (van Casteren & Davis, 2006). For randomization, stimulus sentences were divided into three blocks (A, B and C). In each block, each introducing sentence occurred only once and the second sentence occurred with a different pronoun in each block. So each participant saw each introducing sentence three times, each time with a different pronoun in the second sentence so that the critical manipulation (Pronoun Condition) was within-subjects. Participants were assigned to one of three block orders (ABC, BCA, CAB) and for each participant a different pseudorandomized stimulus file was created. Stimuli were randomized under the following constraints: One constraint was that the same German gender and the same pronoun could appear on no more than five consecutive trials. Another constraint was that the same sentence type (experimental sentence/one of the filler sentences) could appear on maximum four consecutive trials and maximum four correct or incorrect sentences could appear in a row. Furthermore, no more than five sentences of the same congruency/correctness (correct, pseudocongruent, incongruent, incorrect) could appear in a row. In addition, there was a minimum distance for the same introducing sentence (e.g., *This is a bus*.) of 90 trials. Since there were 96 trials in each block, if an introducing sentence had appeared at the very beginning of the block it was possible that it appeared again at the end of the same block, but mostly the same introducing sentence would appear only once per block (e.g., followed by the pronoun *it* in the first block, *he* in the second block, *she* in the third block). There was a minimum distance of two trials between nouns belonging to the same semantic field (e.g., clothes, fruit, furniture) to avoid priming effects.

C-Test

Because English proficiency of German university students usually is quite high, prior to the EEG experiment a screening session was arranged in order to select candidates with low English proficiency. Since there were several problems with the DIALANG-test (cf. section 4.2.3) in the first experiment[7], this time a different test for measuring language proficiency was used. The C-test

[7] There were some concerns regarding the validity of the DIALANG grammar tests as many participants complained that their score was lower than on other tests, e.g. scoring B1 on the DIALANG test while recently having passed a B2 test elsewhere. This could be due to the fact that minor mistakes as spelling errors or a missing diacritical mark may cause an otherwise acceptable answer to be scored as incorrect resulting in a lower test score. It can also happen that DIALANG rejects correct answers because they are not known to the program. Possibly, the three different versions of the grammar test – according to the score in the vocabulary

developed for course placement by the language center of the Humboldt University Berlin seemed adequate for our purposes. The C-test included five short texts selected from newspaper articles of increasing difficulty with parts of words missing that had to be completed. Test scores were also reported on the CEFR scale (Council of Europe, 2001). The test was completed and evaluated online on the internet pages of the language center.

In the screening session subjects were asked some questions about their language learning history in order to double-check for any exclusion criteria regarding English proficiency/learning history already mentioned in the invitation. They were provided with information on the EEG technique as well. The screening session took about 30 minutes and was reimbursed with € 5. When the participants´ level of English matched the required level (A1/A2/B1[8], 0 - 49 points out of 100), they were invited to take part in the EEG experiment and the reimbursement for the screening session was only paid after participation in the EEG experiment.

---

placement test – do not render comparable results, maybe even distorting results due to floor or ceiling effects. Another possible reason for the diverging results of the DIALANG tests and test scores reported by the participants could be that grammar is rarely tested as explicitly as in the DIALANG test but is usually tested as a part of one of the four skills of reading, writing, listening and speaking. Since grammar is often the most difficult aspect of a language for late L2 learners, this could be the reason for so many of the subjects scoring lower than in other tests. In addition, scores on the vocabulary placement test differed widely as guessing was punished and classifying false words as real words can heavily lower the score. Subjects were instructed not to take risks and to mark a word as real only if they were certain, nevertheless, individually different strategies might have affected results. There were also some technical problems because on a few occasions the DIALANG server was down and testing sessions had to be postponed. Therefore, it was decided that it would be preferable to use a test that can also be done offline with a more transparent scoring method.

[8] Two subjects with a low B2 level on the border to level B1 (scores of 50 and 55) were also invited to the EEG session.

In Table 5.4 an overview of the English proficiency scores of the subjects is given:

Level CEFR	Score	Number of subjects
A1	0 - 19	2
A2	20 - 29	2
B1	30 - 49	22
B2	50 - 64	2

Table 5.4 Test scores and levels of European Frame of Reference and number of subjects for each proficiency level.

Language history questionnaire

The participants in the EEG session were also asked to complete a language learning history questionnaire (an Excel table) at home before participating in the EEG session. The questionnaire was similar to the one used in Experiment 1 and contained questions regarding their learning history and current usage of English and other languages. The results of this questionnaire are reported in section 5.2.1 (cf. Table 5.2).

Word translation/intuitive gender assignment task

After completing the experiment, subjects indicated for all of the nouns and adjectives how well they knew the words (on a scale from 0 - 10, not at all - very well), translated them and indicated for the nouns how “male” or “female” they thought the words were (on a scale from 0 - 10, 0 = very male, 10 = very female).

5.2.3 Procedure

Prior to participation subjects read the information about the EEG technique and signed a consent form. After putting on the electrode caps and preparing the electrodes, subjects read the instructions for the experiment on the computer screen. Additionally, participants were explicitly instructed by the experimenter to move as little as possible and not to blink during sentence reading. Next, a practice phase was presented so that subjects could get used to the word-by-word presentation and to practice blinking between trials. Participants sat approximately 90 cm from the screen and the light was dimmed during the experiment. When the practice phase was completed correctly and the subjects indicated that they had no further questions, the experiment was started.

Each trial began with the presentation of a fixation cross for 500 ms. Sentences were presented word by word and each word appeared for 500 ms. During the interstimulus interval of 100 ms a blank screen was presented. After the interstimulus interval at the end of the sentence the question “Correct?” appeared prompting subjects to make a grammaticality judgment by pressing either the green (Yes) or the red (No) button. Position of the response buttons, that is, the response hand was counterbalanced across subjects. The intertrial interval was 1000 ms. Subjects were instructed to use this interval for blinking. Stimuli were presented in Times New Roman, with white letters against a black background and a font size of 22.

The main experiment consisted of 288 trials, presented in 6 blocks of 48 trials each, between which participants were free to take breaks as long as they wished. After each 24 trials (i.e., in the middle of each block) subjects were allowed to take a short self-timed blinking break. After each 48 trials, that is, between blocks, subjects took a longer self-timed break. The experiment took about 40 - 45 minutes. After that, participants did the word translation task. All in all, the EEG session took about 2 to 2.5 hours, including hair washing.

Words were presented centered on an Acer AL1923 19 inch LCD-monitor with a data resolution of 1024 by 768 pixels. The display device was an NVIDIA Quadro NVS 210S with a display memory of 256.0 MB. The experimental software Presentation ® (version 14.7) by Neurobehavioral Systems, Inc. running on a Dell OptiPlex 740 computer with an AMD Athlon(tm) 64 X2 Dual Core Processor 5600+and Windows XP Professional operating system was used for stimulus presentation. ERPs were recorded using Brain Vision Recorder (version 1.03.0004) by Brain Products GmbH also running on a Dell OptiPlex 740 with an AMD Athlon(tm) 64 X2 Dual Core Processor 5600+ and Windows XP Professional operating system. Signals were amplified with Brain Vision BrainAmp DC (32-channel-EEGamplifier), connected to the computer with the Brain Vision USB2 Adapter. An Easy-Cap Electrode Input Box EiB32 was used.

5.2.4 EEG recording and data analysis

Scalp voltages were collected from 25 Ag/AgCl sintered ring electrodes mounted on an elastic cap (Modular EEG Recording Cap by EASY CAP) and distributed symmetrically over the left and right hemispheres according to the international 10–20 system. Electrode sites (F3, F7, FC5, Fz, FCz, F4, F8, FC6, T7, C3, Cz, C4, T8, CP5, P7, P3, Pz, CPz, POz, CP6, P8, P4, O1, Oz, O2) covered frontal, temporal, parietal, and occipital areas. See Figure 5.1 for a diagram of the electrode positions used. Impedances were kept below 5 kΩ. EEG signals were sampled at 250 Hz, amplified and filtered during recording with a low cut-off filter of ca. 0.02 Hz and a high cut-off 1000 Hz. An electrode near the Cz electrode was used as the ground electrode and all electrodes were referenced to linked mastoids. Reference electrodes were placed on both mastoids. Eye movements and blinks, that is, horizontal and vertical electrooculograms (EOGs) were measured with supra- and infraorbital electrodes at the right eye and with electrodes in the outer canthi of both eyes.

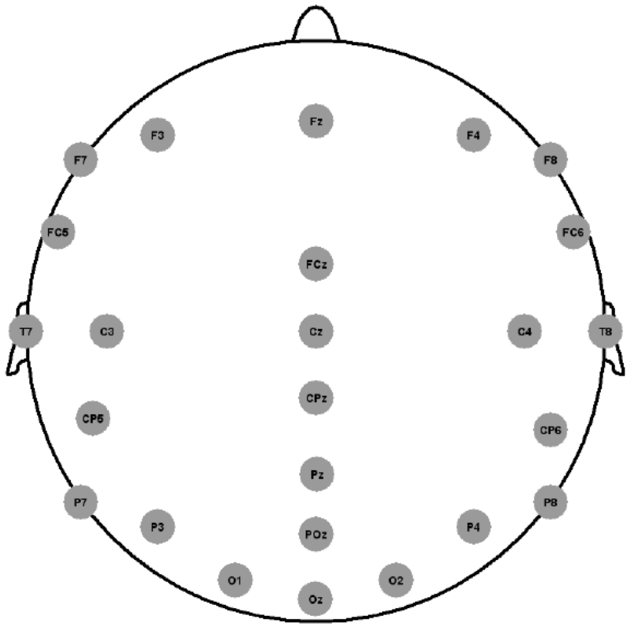


Figure 5.1 Electrode placement for Experiment 2[9].

EEG data were processed and exported with Brain Vision Analyzer Software. Prior to data analysis the data were filtered with a low cut-off filter of 0.02 Hz and a high cut-off filter of 28 Hz. Epochs had a length of 1400 ms, starting 200 ms prior to the onset of the critical pronoun and continuing till 1200 ms thereafter. Trials were corrected to a baseline of 200 ms before stimulus onset[10]. Artifacts were rejected in semiautomatic mode according to the following criteria: maximum voltage step of the gradient 50 µV/ms, maximum difference of values in intervals 200 µV, amplitude between -100 µV and 100 µV and the lowest allowed activity 1 µV. Less than 0.5 % of the data per subject had to be rejected due to artifacts. Blink detection and ocular correction was achieved with the algorithm by Gratton, Coles and Donchin (1983). Overall, subjects were quite successful in avoiding blinking during trials. The mean number of blinks in the 144 segments (the experimental sentences) that entered data analysis was 6.71 (4.66 %). Next, average ERPs per subject per condition and grand averages over subjects were calculated and mean area voltages (µV * ms) were exported for statistical analysis. The main component of interest was the P600 component, corresponding to a time window of 500 - 800 ms and time-locked to the personal pronouns of the second sentence. Further components (N400, P200) time-locked to the personal pronoun or time-locked to subsequent items were analyzed if visual inspection gave reason for this, such as a late negativity which occurred after the verb *is*.

For the time windows in question, mean area voltages were subjected to repeated measures analyses of variance (General Linear Model; GLM). The analysis was conducted in a hierarchical manner. First, the GLM was carried out with the factors Pronoun Condition and Region of Interest (ROI). Six ROIs were created: posterior midline (CPz, Pz, POz), posterior right (CP6, P8, P4), posterior left (CP5, P7, P3), anterior midline (Fz, FCz, Cz), anterior right (F4, F8, FC6, C4), and anterior left (F3, F7, FC5, C3). Temporal and occipital electrodes were not included in the analysis. Greenhouse-

[9] Figure adapted with kind permission from Juliane Domke, Psycholinguistic EEG laboratory, Humboldt University of Berlin.

[10] Since the interstimulus interval was 100 ms, the baseline also contains the last 100 ms of the presentation time of the last word of the previous sentence. Since the introducing sentences are the same across conditions, no differences between the conditions should arise from that.

Geisser corrections for sphericity violations were always applied as suggested by Luck (2005, p. 258). For ease of reading, reported degrees of freedom are uncorrected but *p*-values are corrected. If an interaction of Pronoun Condition with a ROI relevant for the component in question was significant, it was followed up by an analysis of the effect of Pronoun Condition within each ROI. After that, pairwise comparisons between conditions were carried out with *t*-tests for the ROIs with significant effects of Pronoun Condition. Because my hypotheses for the P600 (time-locked to the pronouns *he*/*she*/*it*) were directional, *t*-tests for this component were one-tailed. Since I had no a priori hypotheses regarding the observed P200 (also time-locked to the pronouns *he*/*she*/*it)* and the sustained negativity (time-locked to the verb *was* following the pronouns), *t*-tests for those components were two-tailed. Bonferroni-corrections for multiple comparisons were applied. In all figures depicting EEG activity, negative was plotted upward.

For the overall analyses of the ERP components, the results of two analyses are reported: one analysis including only correct trials (and only those 17 subjects with at least 20 trials per condition) and a second analysis including both correct and incorrect trials. For the subsequent analyses investigating proficiency effects only the latter analysis was reported because in the former analysis most of the subjects with lower proficiency had to be excluded making it impossible to investigate proficiency effects in this manner.

Moreover, the analysis including correct as well as incorrect trials also seemed necessary in order to compensate for possible distorting effects due to small and/or unequal numbers of trials in some experimental conditions and proficiency groups after removal of error trials. As we will see in the next section, the number of rejected trials differed systematically between experimental conditions and groups, because significantly more errors were made in the pseudocongruent condition and by the low-proficient group. Especially the exclusion of incorrect trials in the pseudocongruent condition seemed problematic for the purpose of this study because the incorrect trials of the pseudocongruent condition actually are the most interesting trials, that is, the trials where gender transfer was supposed to occur. For these same reasons the third possible option, an analysis excluding incorrect trials but including all subjects, was discarded (but was in some cases reported in a footnote). Carrying out the analysis including correct as well as incorrect trials seemed to be the most suitable approach for investigating the present research questions.

Analysis behavioral results

Error rates were analyzed with a GLM with repeated measures. The first analysis was carried out with the factor Pronoun Condition (correct, incongruent, pseudocongruent) and the second analysis with the factors Pronoun Condition and Level (low vs. high-proficient, as measured by the C-test). If there was a significant interaction of Level with Pronoun Condition, pairwise comparisons between conditions were carried out with *t*-tests.

RTs were not analyzed because inspection of the descriptive data suggested that the RT results were severely distorted due to the fact that the response for the GJT had to be made only at the end of the sentence, 2500 ms after the actual error had occurred. On about 15 % of the trials, RTs were either extremely low (< 250 ms) or extremely high (> 2500 ms). This diverse response pattern suggests that in many cases subjects either initiated the response process before the actual question appeared on the screen, or had difficulties remembering the required response at the end of the sentence. For these reasons, I refrained from conducting an RT analysis.

In the following section, I will first present and discuss the behavioral results, namely analysis of the error rates in the GJT, before turning to the ERP data.

5.3 Behavioral results

First, the general error rates across conditions were analyzed. Then, possible effects of language proficiency (as measured by the C-test) were investigated. On the whole, the error rate for the pseudocongruent condition was much higher than for the other two conditions. The incongruent condition proved to be the easiest condition with the least amount of errors (cf. Table 5.5 and Figure 5.2).

	Correct	Incorrect	% incorrect
Correct	1,004	340	25.3%
Pseudocongruent	844	500	37.2%
Incongruent	1,183	161	12.0%

Table 5.5 Amount of absolute correct and incorrect grammatical decisions, and percentage of incorrect decisions, across Pronoun Conditions.

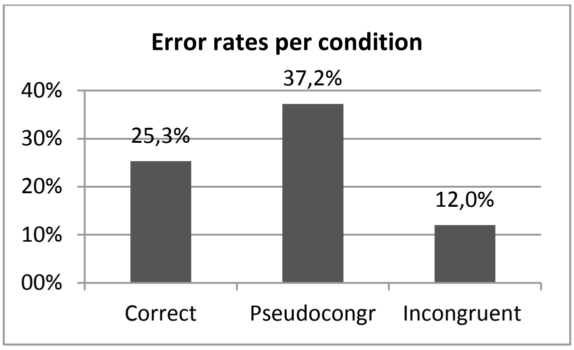


Figure 5.2 Error rates per Pronoun condition in percent.

The statistical analysis confirmed that there was indeed a significant main effect of Pronoun Condition ($F_1(2,54) = 10.824$, $p < 0.001$, $F_2(2,94)$, $= 132.320$, $p < 0.001$). Pairwise comparisons with *t*-tests for the subject and item-analysis revealed that all the differences between the conditions were significant (cf. Table 5.6).

	t1			*t2*		
	df	*t*	*p*	*df*	*t*	*p*
Correct – Pseudocongruent	27	2.699	.012*	47	7.99	< .001*
Correct – Incongruent	27	- 2.286	.030*	47	- 8.449	< .001*
Pseudocongruent – Incongruent	27	- 4.273	< .001*	47	- 15.905	< .001*

Table 5.6 Results of pairwise comparisons of Pronoun conditions with *t*-tests. Significant results are marked with an asterisk.

Effects of language proficiency

In order to investigate possible effects of language proficiency, subjects were divided into two groups, a high-proficient and a low-proficient group with 14 subjects in each group, by median split according to their score on the C-test. When applying a median split to the data, four subjects with the same test score (43 points) had to be arbitrarily assigned to one of the groups. Nevertheless, the difference in proficiency between the two groups was significant ($t = 4.520$, df = 26, $p < 0.001$).

An overview of the proficiency scores of the two proficiency groups is given in Table 5.7.

	Language proficiency C-Test	
	Mean (SD)	**Range**
Low-proficient group	34.79 (9.41)	15 - 43
High-proficient group	46.79 (3.19)	43 - 55

Table 5.7 The mean score, standard deviation (in parentheses), and range of the scores of the low-proficient and high-proficient group on the C-test.

An overview of the error rates per proficiency group is given in Table 5.8 and Figure 5.3.

		Correct	Incorrect	% incorrect
Low-proficient group	**Correct**	548	124	18.5 %
	Pseudo-congruent	513	159	23.7 %
	Incongruent	637	35	5.2 %
High-proficient group	**Correct**	456	216	32.1 %
	Pseudo-congruent	331	341	50.7 %
	Incongruent	546	126	18.8 %

Table 5.8 Amount of absolute correct and incorrect grammatical decisions and percentage of incorrect decisions, across Pronoun Conditions and proficiency groups according to C-test.

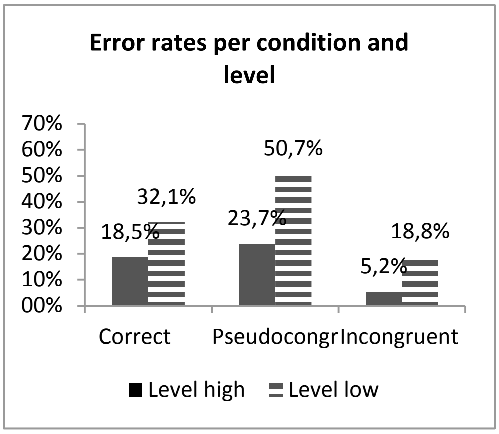


Figure 5.3 Error rates per condition and level in percent.

As can be seen in Figure 5.3, the error pattern across conditions was the same for both proficiency groups: The error rate for the pseudocongruent condition was much higher than for the other two conditions. As before, the incongruent condition seemed to be the easiest condition with the least amount of errors in both proficiency groups. This pattern matches the result of the statistical analysis where I found a main effect of Pronoun Condition ($F_1(2, 52) = 10.836$, < 0.001; $F_2(2, 141) = 145.083$, $p < 0.001$). Overall, low-proficient subjects made more errors than high-proficient subjects (34 % vs. 16 %), which proved to be significant ($F_1(1, 26) = 4.192$, $p = 0.051$; $F_2(1, 141) = 319.680$, $p < 0.001$). The interaction effect of Pronoun Condition with Level was significant for F_2 ($F_1(2, 52) = 1.029$, $p = 0.357$; $F_2(2, 141) = 19.655$, $p < 0.001$). Follow-up analyses revealed that the effect of Level was significant in the high-proficient ($F_1(2,26) = 5.315$, $p = 0.030$, $F_2(2,141)$, = 59.007, $p < 0.001$) as well as the low-proficient group ($F_1(2,26) = 6.184$, $p = 0.006$, $F_2(2,141)$, = 117.391, $p < 0.001$). Results of pairwise comparisons with *t*-tests were also similar between proficiency groups (cf. Table 5.9). In the analysis across participants, only the difference between the pseudocongruent and incongruent condition was significant for both proficiency groups, while in the analysis across items all three comparisons between conditions were significant.

	***t*1**	***t*2**

		df	t	p	df	t	p
Low-proficient group	**Correct – Pseudocongruent**	13	2.311	.057	94	8.916	< .001*
	Correct – Incongruent	13	- 1.305	.321	94	- 6.042	< .001*
	Pseudocongruent – Incongruent	13	- 3.568	.005*	94	- 16.042	< .001*
High-proficient group	**Correct – Pseudocongruent**	13	1.722	.164	94	2.809	.009
	Correct – Incongruent	13	- 2.215	.068	94	- 7.576	< .001*
	Pseudocongruent – Incongruent	13	- 2.438	.045*	94	- 11.211	< .001*

Table 5.9 Results of pairwise comparisons of conditions with *t*-tests for the low-proficient and high-proficient group. Significant results are marked with an asterisk.

Summary

Error rates showed that, as expected, the pseudocongruent condition was more difficult than the incongruent condition, reflected in significantly higher error rates in the GJT. So there was a clear sign of L1 gender transfer in the error rates. Also as predicted, the incongruent condition was the easiest condition with the least mistakes, even less than in the correct condition. This can be explained by the fact that only the incongruent condition can be thought of as completely free of L1 gender transfer as the incongruent condition, contrary to the correct condition, would be incorrect in both languages. The correct condition, in contrast, was correct in English but would be incorrect for the German translation. In the analysis investigating the effects of language proficiency, overall error rates were higher for the low-proficient group than for the high-proficient group but the gender transfer effect in error rates was equally strong for both proficiency groups.

5.4 ERP results

First, I will summarize the analyses of the main component of interest, the P600 component[11]. Then, I will discuss the analysis of two additional components that were not predicted, a P200 and a sustained negativity. The P200 was observed time-locked to the pronouns *he*/*she*/*it* and the sustained negativity component occurred after the verb *was* following the pronouns.

5.4.1 P600 component

For the time window from 500 - 800 ms, mean area voltages were analyzed with a GLM with repeated measures. As mentioned in section 5.2.4, the results of two analyses are reported, one analysis including only correct trials and only subjects with at least 20 trials per condition (n = 17) and a second analysis including all 28 subjects and both correct and incorrect trials. In the subsequent analyses, I will look at the effects of language proficiency by splitting up subjects into two groups, first by language proficiency as measured by the C-test and then according to the pattern of error rates observed in the behavioral analysis. These analyses were carried out with correct and incorrect trials included, for the reasons explained earlier.

[11] In the analysis of proficiency according to error pattern visual inspection gave reason to an additional analysis for the N400 time window for a subset of the participants.

Overall Analysis

Only correct trials

Visual inspection of the ERP patterns (cf. Figure 5.4) showed a substantial difference in mean area voltages between the pseudocongruent and incongruent condition in comparison with the correct condition in the P600 time window at posterior electrode sites.

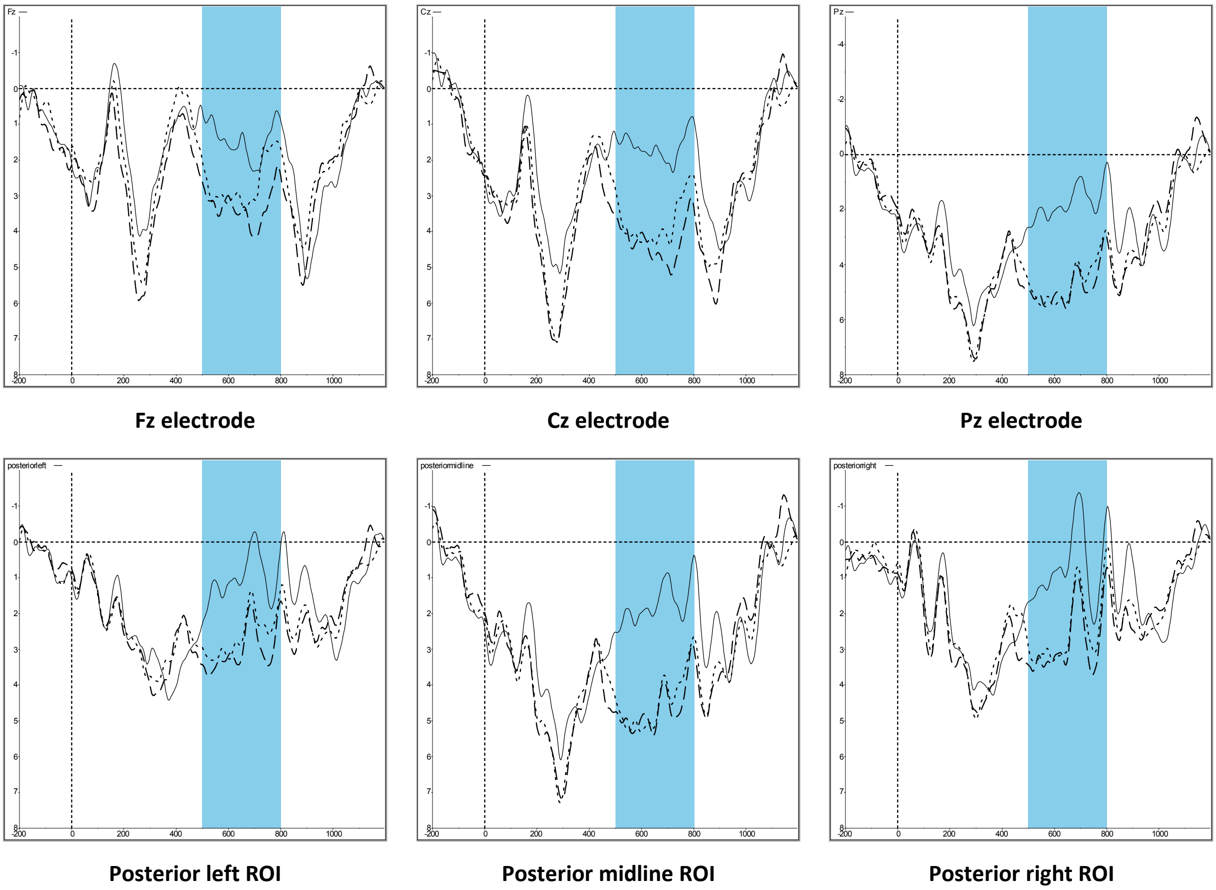


Figure 5.4 Grand Average waveforms of 17 subjects and only correct trials, shown on midline Fz, Pz, and Cz electrodes and posterior midline, posterior right, and posterior left ROIs. Continuous line = correct condition, dashed line = incongruent condition, dotted line = pseudocongruent condition.

Contrary to my expectations, though, at least at face value there seemed to be no difference between the incongruent and pseudocongruent condition. At anterior electrode sites there were no obvious differences between the conditions.

The results of the statistical analysis are displayed in Table 5.10. As the interaction of ROI with Pronoun Condition was significant, next, an analysis of the effects of Pronoun Condition within each ROI was carried out

.

	df	*F*	*p*
ROI	5, 80	7.906	< .001*
Pronoun Condition	2, 32	3.746	.040*
ROI * Pronoun Condition	10, 160	2.852	.032*

T**able 5.10** Results of the GLM with the factors ROI and Pronoun Condition. Significant results are marked with an asterisk.

		df	*F*	*p*
ROIs	**Posterior midline**	2, 32	5.667	.015*
	Posterior right	2, 32	2.540	.106
	Posterior left	2, 32	4.713	.018*
	All anterior			> .05

Table 5.11 The effects of Pronoun Condition within each ROI. Significant results are marked with an asterisk.

As can be seen in Table 5.11, Pronoun Condition had a significant effect in all posterior ROIs except for the posterior right ROI. So in the next step, pairwise comparisons with *t*-tests were conducted only for the posterior midline and left ROIs.

		df	*t*	*p*
Posterior midline ROI	**Correct – Pseudocongruent**	16	- 2.646	.054
	Correct – Incongruent	16	- 2.504	.069
	Pseudocongruent – Incongruent	16	- 0.413	1.000
Posterior left ROI	**Correct – Pseudocongruent**	16	- 2.743	.042*
	Correct – Incongruent	16	- 1.714	.318
	Pseudocongruent – Incongruent	16	- 1.568	.408

Table 5.12 Results of pairwise comparisons with *t*-tests (one-tailed, Bonferroni-corrected) for posterior midline and posterior left ROI. Significant results are marked with an asterisk.

For the posterior midline ROI, *t*-tests showed that the difference between the correct and incongruent condition was marginally significant. The difference between the correct and pseudocongruent condition was almost significant, too. For the posterior left ROI, only the difference between the correct and pseudocongruent condition was significant. Hence, consistent with the visual inspection pattern, no significant differences between the incongruent and pseudocongruent condition were found. An additional analysis was carried out with only correct trials but all subjects included. Overall, the results were similar but the differences between the correct and incongruent condition were also significant[12].

[12] Analysis only correct trials: At the posterior midline ROI there was a significant difference between the correct and incongruent condition ($t(27) = -2.615$, $p = 0.040$) and between the correct and pseudocongruent condition ($t(27) = -2.789$, $p = 0.030$). At the posterior right ROI there was a significant difference between the correct and incongruent condition ($t(27) = -2.621$, $p = 0.040$).

Correct and incorrect trials

The results of the analyses were similar as in the previous analysis. In the first analysis, the interaction of ROI and Pronoun Condition was also (almost) significant ($F(10, 270) = 2.292$, $p = 0.051$). In the following step, there was also a main effect of Pronoun Condition only at the posterior midline ($F(2, 54) = 3.862$, $p = 0.033$) and posterior left ROI ($F(2, 54) = 3.463$, $p = 0.039$). As before, there were no significant results in any of the anterior ROIs as the P600 component had a clear posterior distribution. The results of the *t*-tests were slightly different with the difference between the correct and pseudocongruent condition reaching only marginal significance at the posterior midline ($t(27) =-2.348$, $p = 0.078$) and posterior left ROI ($t(27) =-2.105$, $p = 0.072$).

Summary

The analysis with 17 subjects and only correct trials confirmed the pattern observed upon visual inspection. The difference between the correct condition and the two grammatically incorrect conditions (the incongruent and the pseudocongruent condition) was significant at posterior ROIs. In the analysis with all subjects and correct as well as incorrect trials included, only the difference between the correct and pseudocongruent condition reached significance at posterior ROIs. There were no significant results in any of the anterior ROIs as the P600 component had a clear posterior distribution. This is consistent with the expected P600 distribution according to the literature (e.g., Luck, 2005).

Contrary to my expectations and different from the results of the error rates, no significant differences between the incongruent and pseudocongruent condition were found. For some reason, the huge difference in error rates between the incongruent and pseudocongruent condition shown in Table 5.5 was not reflected in the P600 component because an equally strong P600 was observed for both incorrect conditions compared to the correct condition. This was a pattern that would be expected for native English speakers or very high-proficient L2 speakers. It is surprising that even the, on the whole, very low-proficient participants of the present experiment already showed such a processing pattern. In the following analyses, subjects were divided into two groups according to proficiency. In one analysis, subjects were divided into two proficiency groups according to their C-test scores, in another analysis subjects were divided into two proficiency groups according to whether they had shown strong gender transfer in error rates or not. One goal of these analyses was to investigate proficiency effects. A second aim was to test whether L1 gender transfer effects were, in fact, present in the ERP data (of low-proficient subjects).

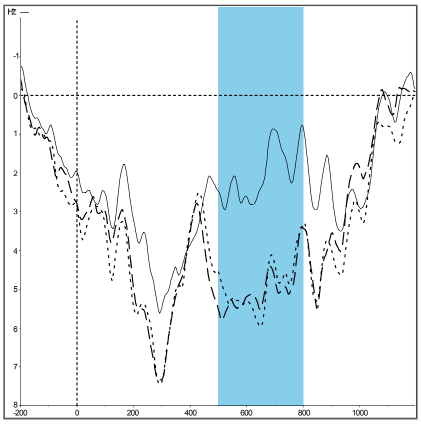
Next, I will discuss the effects of language proficiency as measured by the C-test and in a subsequent analysis the effects of language proficiency as indicated by the pattern of error rates. Because of the reasons mentioned above, these analyses were conducted with correct and incorrect trials included.

Effects of language proficiency

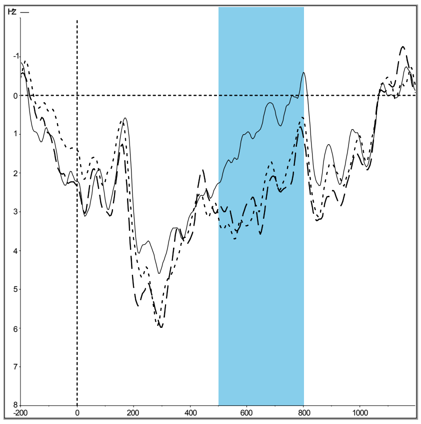
Language proficiency according to the C-test

In order to investigate possible effects of language proficiency, in this analysis subjects were divided into two groups as in the behavioral analysis above, namely, a high-proficient and a low-proficient group according to their score on the C-test. In the behavioral analysis, there was a significant difference in error rates for the two levels.

Upon visual inspection, the pattern seemed to be different for the two groups, with the high-proficient group showing a P600 pattern for both incorrect conditions, as in the overall analysis. The low-proficient group also seemed to show a P600, but it was much smaller. In the statistical analysis, however, there was no significant interaction of ROI, Pronoun Condition, and Level ($F(10, 260) = 0.262$, $p = 0.926$)[13]. Hence, the perceived difference between the two proficiency groups at face value was spurious. The effect of Level was marginally significant ($F(1, 26) = 3.592$, $p = 0.069$). In the analysis including only correct trials but all subjects, the result was similar[14].



Pz electrode of the high-proficient group



Pz electrode of the low-proficient group

Figure 5.5 Grand Average waveforms of the high-proficient and low-proficient group, shown on Pz electrode. Continuous line = correct condition, dashed line = incongruent condition, dotted line = pseudocongruent condition.

Summary

When subjects were divided into two groups according to C-test scores, no difference between the high-proficient and low-proficient group was found. Upon visual inspection, the ERP pattern for both groups seemed to be different with the less proficient group exhibiting a smaller P600, but the difference in mean area voltage was not significant. This might be due to the fact that, as most of the subjects had a score between 30 and 49 (level B1; 22 subjects), the two groups were still very similar regarding their test scores. Since even our overall quite low-proficient language learners already showed a processing pattern that would be expected for native speakers, maybe including more very

[13] Note that the lack of finding an interaction of ROI, Pronoun Condition, and Level cannot be attributed to a lack of discriminatory power between the two proficiency groups. In an analysis excluding the four subjects with overlapping C-test scores leaving 12 subjects in each group, the relevant interaction of ROI, Pronoun Condition and Level was not significant, either ($F(10, 220) = 0.659$, $p = 0.720$). But there was a main effect of Level ($F(1, 22) = 5.991$, $p = 0.023$).

[14] Analysis only correct trials: The relevant interaction of ROI, Pronoun Condition, and Level was not significant $F(10, 260) = 0.537$, $p = 0.762$ and neither was there a main effect of Level $F(1, 26) = 1.018$, $p = 0.322$.

low-proficient learners with very low C-test scores (e.g., level A2) would have rendered discriminatory power stronger leading to a significant difference between the two groups. Nevertheless, it is also possible that the C-test is just not a fine-grained enough measure to be able to make predictions about such subtle processing differences. For the next analysis of proficiency effects a proficiency criterion more related to the experimental task was used.

Language proficiency according to error pattern

In the analysis just reported, there were no significant differences between the two groups, even though visually, the ERP patterns looked quite different. Therefore, I tried to find a more decisive criterion for dividing subjects into two proficiency groups. In addition, since the results obtained in the behavioral analysis are not reflected in the ERP analysis, I wanted to find out if an effect of L1 gender transfer was present in the ERP results for a subset of the participants. It was also possible that in the ERP analysis, contrary to the behavioral analysis, potential gender transfer effects were covered up by (a majority of) subjects lacking gender transfer. Thus, in order to investigate how gender transfer effects would be reflected in the ERP results, this time, subjects were divided into two groups according to whether they showed strong gender transfer in the error rates or not[15]. In the “L1 transfer group” those participants who had at least 10 % more erroneous responses in the pseudocongruent than in the incongruent condition were included. The “no L1 transfer group” consisted of participants whose difference in error rate between those two conditions was less than 10 %. Based on this criterion, 13 subjects were assigned to the “L1 transfer group” and 15 subjects were assigned to the “no L1 transfer group”. Figure 5.6 shows the considerable difference in gender transfer for the two groups. The “L1 transfer group” exhibited big differences in the error rates between conditions, while the error rates in the “no L1 transfer group” were very similar across conditions. By and large, the ”L1 transfer group” also made a lot more errors (48 %) and seemed less proficient than the “no L1 transfer group” (5 % errors), at least with respect to the experimental task. Yet, in spite of those differences regarding L1 gender transfer and overall error rates, no big proficiency differences concerning their general knowledge of English were found in the metadata. Mean C-test scores for the two groups were 43.47 (SD 8.49) for the “no L1 transfer group” and 37.69 (SD 8.67) for the “L1 transfer group”. Mean self-assessment scores of English proficiency (scores of reading, speaking and writing combined) were 4.09 (SD 0.97) and 3.69 (SD 1.16), respectively. Thus, the two proficiency groups were different concerning their processing of gender and overall error rates but less so regarding their general knowledge of English, as evidenced by C-test scores and other metadata.

[15] For a similar approach dividing subjects into two proficiency groups according to error rates (“fast learners” and “slow learners”) for the ERP analysis, compare Osterhout et al. (2006, p. 217).

		Correct	Incorrect	% incorrect
No L1 Transfer group	Correct	684	36	5.0 %
	Pseudo-congruent	674	46	6.4 %
	Incongruent	692	28	3. 9 %
L1 transfer group	Correct	320	304	48.7%
	Pseudo-congruent	170	454	72.8%
	Incongruent	491	133	21.3%

Table 5.13 Amount of absolute correct and incorrect grammatical decisions, and percentage of incorrect decisions, across Pronoun Conditions for the "no L1 transfer group" and the "L1 transfer group".

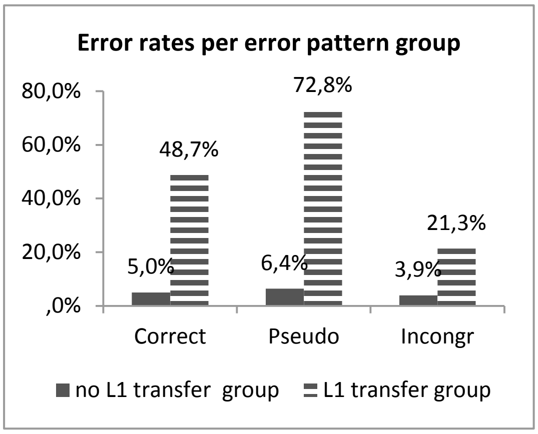
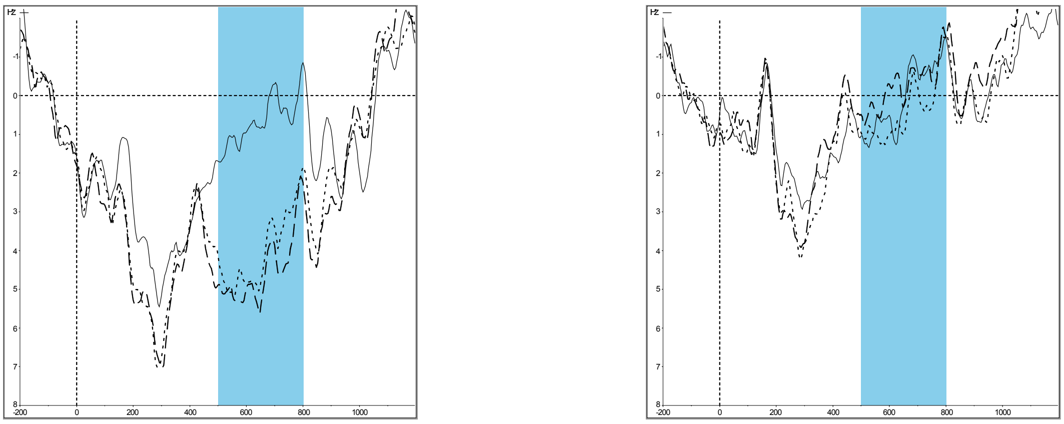


Figure 5.6 Error rates across conditions for the "no L1 transfer group" and the "L1 transfer group". Overall error rates are 5 % for the "no L1 transfer group" and 48 % for the "L1 transfer group".

The observed ERP patterns of the two groups can be seen in Figure 5.7. Upon visual inspection, there seemed to be a quite substantial difference in processing between the two groups. The "no L1 transfer group" showed the effect obtained in the overall analysis, that is, a clear difference between the correct and incorrect conditions in the P600 time window. The "L1 transfer group", on the other hand, showed a different pattern. Even though the three conditions seemed to differ somewhat from each other, with the waveform of the incorrect conditions being somewhat more positive than the waveform of the correct condition, no strong P600 effect was visible and neither was it clear if there was an effect of Pronoun Condition. On the contrary, it seemed as if there was a negativity for the two incorrect conditions just before the P600 time window, which could be an N400 effect and will also be analyzed.



Pz electrode of the "no L1 transfer group" **Pz electrode of the "L1 transfer group"**

Figure 5.7 Grand Average waveforms of the "no L1 transfer group" and the "L1 transfer group," shown on the Pz electrode. The "no L1 transfer group" did not show the expected pattern in the error rates. The "L1 transfer group" showed the expected pattern in the error rates. Continuous line = correct condition, dashed line = incongruent condition, dotted line = pseudocongruent condition.

As can be seen in Table 5.14, there was a significant interaction of ROI with Pronoun Condition and Transfer Group ("no L1 transfer group" vs. "L1 transfer group").[16]

	df	*F*	*p*
ROI	5, 130	7.913	< .001*
ROI * Transfer Group	5, 130	2.475	.067*
Pronoun Condition	2, 52	1.963	.155
Pronoun Condition * Transfer Group	2, 52	1.870	.168
ROI * Pronoun Condition	10, 260	2.155	.062*
ROI * Pronoun Condition * Transfer Group	10, 260	2.574	.029*
Transfer Group	1, 26	1.717	.202

Table 5.14 Results of the GLM with repeated measures factors ROI, Pronoun Condition, and Transfer Group ("no L1 transfer group" vs. "L1 transfer group"). Significant results are marked with an asterisk

When the subsequent analysis was split up according to Transfer Group, there was a significant interaction of ROI and Pronoun Condition for the "no L1 transfer group" (cf. Table 5.15). However, there were no significant results for the "L1 transfer group" (cf. Table 5.16). As a consequence, only the analysis for the "no L1 transfer group" was continued.

	df	*F*	*p*
ROI	5, 70	8.738	< .001*
Pronoun Condition	2, 28	3.715	.050*
ROI * Pronoun Condition	10, 140	3.641	.011*

Table 5.15 Results of the GLM for the "no L1 transfer group" with the factors ROI and Pronoun Condition. Significant results are marked with an asterisk.

	df	*F*	*p*
ROI	5, 60	1.394	.264
Pronoun Condition	2, 24	.027	.973
ROI * Pronoun Condition	10, 120	.799	.523

Table 5.16 Results of the GLM for the "L1 transfer group" with the factors ROI and Pronoun Condition. Significant results are marked with an asterisk.

The analyses of the effects of Pronoun Condition within each ROI revealed significant effects for the posterior midline ($F(2, 28) = 7.098$, $p = 0.008$) and posterior left ROI ($F(2, 28) = 5.644$, $p = 0.013$). Pairwise comparisons with *t*-tests (Table 5.17) showed that at the posterior midline ROI there was a significant difference between the mean area voltage of the correct and incongruent condition and between the correct and pseudocongruent condition. At the posterior left ROI, there was a significant difference between the correct and pseudocongruent condition.

[16] In the analysis with only correct trials, the interaction of ROI with Pronoun Condition and Level was not significant ($F(10, 260) = 0.537$, $p = 0.720$). This was possibly due to the fact that there were not enough trials left in the less proficient group for any differences to become significant (only 981 of 1872 trials, i.e., 52 %). Since the groups were divided by error rate, the analysis of one of the groups by definition contains more incorrect and hence less trials than the other group, possibly skewing the results. So in this case, the analysis based on correct and incorrect trials seemed definitely more reliable.

		df	t	p
Posterior midline ROI	**Correct – Pseudocongruent**	14	- 2.936	.033*
	Correct – Incongruent	14	- 2.796	.042*
	Pseudocongruent – Incongruent	14	- 0.633	1.000
Posterior left ROI	**Correct – Pseudocongruent**	14	- 2.846	.039*
	Correct – Incongruent	14	- 1.745	.309
	Pseudocongruent – Incongruent	14	- 2.026	.186

Table 5.17 Results of pairwise comparisons with t-tests for the “no L1 transfer group”. Significant results are marked with an asterisk.

N400

As mentioned above, for the L1 transfer group, the ERP patterns for the two incorrect conditions seemed to be more negative than for the correct condition in the N400 time window (350 - 550 ms; cf. Table 5.9). Since, as mentioned in section 5.1, it is known from the literature that sometimes low-proficient learners relying on surface strategies show an N400 instead of a P600 when processing grammatical errors (Guo, Guo, Yan, Jiang, & Peng, 2009; McLaughlin, Tanner, Frenck-Mestre, Valentine, & Osterhout, 2010; Morgan-Short, Steinhauer, Sanz, & Ullman, 2012; Osterhout, McLaughlin, Pitkänen, Frenck-Mestre, & Molinaro, 2006, experiment 2; Steinhauer, White, & Drury, 2009), this potential N400 effect was also investigated. However, the three-way interaction of ROI with Pronoun Condition and Transfer Group ($F(10, 260) = 1.379$, $p = 0.245$), was not significant so the analysis was not continued.

Summary

The aims of this analysis were to investigate proficiency effects and to investigate whether the gender transfer effect found in the error rates would also be present in the ERP data for a subset of the participants. To this end, subjects were divided into two groups, according to whether they showed gender transfer in the behavioral data or not. Upon visual inspection, though, even for the “L1 transfer group” no gender transfer was visible in the ERPs and neither was there a P600 for any of the conditions. Apparently, the “L1 transfer group” showed no sensitivity to grammatical violations in the English sentences. The “no L1 transfer group”, however, showed the same pattern as in the overall analysis, that is, a P600 for both incorrect conditions compared to the correct condition. These observations were confirmed by statistical analyses. Moreover, upon visual inspection, it seemed as if there was an N400 effect for the incorrect conditions for the “L1 transfer group”, but the difference between mean area voltages was not significant.

Discussion behavioral results and P600

Behavioral results: The pattern of error rates in the GJT was as predicted, with higher error rates for the pseudocongruent condition than for the incongruent condition, due to gender transfer from the L1. The incongruent condition, on the other hand, proved to be the easiest condition with the lowest error rate. This is probably due to the fact that this condition is grammatically incorrect in both L1 and L2, so that no gender transfer from the L1 could arise. The L2 sentences in the correct condition, in contrast, are more difficult to process from a transfer perspective. If translated literally into the L1, the pronoun *it* would be incorrect, possibly also leading to small negative L1 gender transfer effects. The low-proficient group (as measured by the C-test) made more mistakes than the high-proficient group, but L1 gender transfer was equally strong for the two groups.

P600: The gender transfer effect observed in the behavioral data was not reflected in online processing. A P600 effect was expected for the incongruent condition and in a weaker form for the pseudocongruent condition. However, no difference between the pseudocongruent and incongruent condition was observed. There was an equally pronounced P600 for both incorrect conditions as compared to the correct condition. This result, demonstrating that subjects differentiated between syntactically correct and incorrect conditions without showing L1 transfer, was a result that would have been expected for native speakers or very proficient L2 speakers. It was surprising that even the overall very low-proficient participants of the present experiment showed such a processing pattern. When subjects were divided into two groups according to language proficiency as measured by the C-test, no difference between the two groups was found.

In order to further investigate proficiency effects and the discrepancy between behavioral and online results, that is, the missing L1 gender transfer effect in the ERP results, subjects were divided into two groups according to whether they exhibited gender transfer at the behavioral level or not. The two groups did indeed show different processing patterns, albeit not quite as expected. For the group with gender transfer in the behavioral data, there was no evidence of gender transfer in the ERP data. What is more, this group did not even show sensitivity to grammatical errors in the English sentences at all. The group revealing no gender transfer in the error rates, however, showed a P600 component in response to both incorrect conditions, that is, the effect we also saw in the overall analysis and that would be expected for native or very high-proficient L2 speakers.

From these data, it appears that even our low-proficient speakers, who seemed quite homogeneous in a lot of aspects, can be divided into two groups, one showing no sensitivity to grammatical violations and the other one behaving like native speakers. Surprisingly, at least in these online processing results there was no evidence for an (intermediate) stage where L2 speakers experience L1 transfer. It rather appeared as if these beginning speakers of English “jumped” from a very low-proficient stage without any sensitivity to L2 grammatical violations to the highest stage with a processing pattern like native speakers. A potential explanation could be that the structure I tested was probably a structure that is fairly easy to acquire, therefore possibly allowing learners to skip a stage and progress from apparently no grammatical sensitivity to the highest, almost native-like stage. This would be in line with studies discussed in section 1.4.2 which have reported L2 learners progressing from no L2 knowledge to almost native-like ERP patterns within few hours of instruction (e.g., Davidson & Indefrey, 2009; Friederici, Steinhauer, & Pfeifer, 2002; Mueller, Hahne, Fuji, & Friederici, 2005). Another possibility is that I failed to find evidence for an “intermediate stage” due to a lack of subjects in the ”L1 transfer group”.

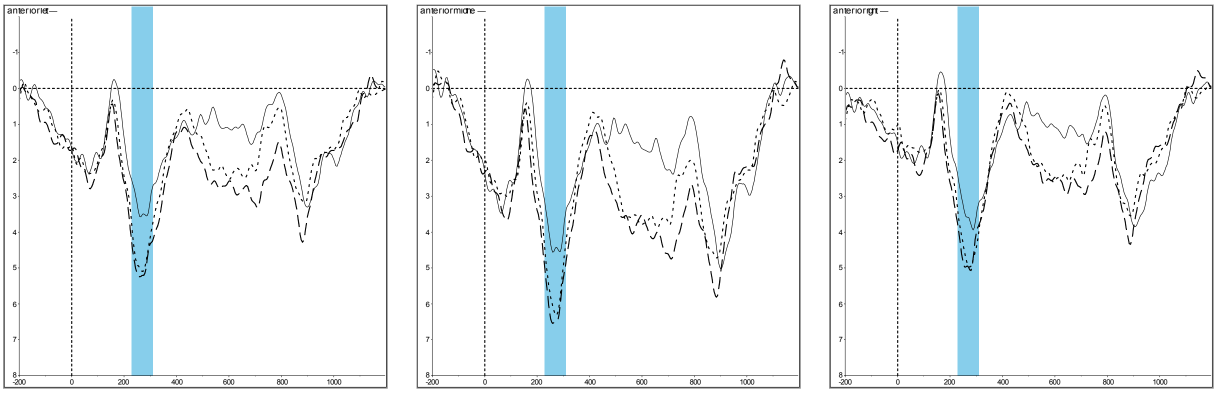
5.4.2 P200 component

There were no a priori hypotheses regarding the P200 component, but upon visual inspection differences between conditions were found in this time window so analyses of this component are reported.

Mean area voltages in the time window from 230 - 310 ms were analyzed with a GLM with repeated measures. As mentioned in section 5.2.4, for the overall analyses, the results of two analyses are reported, one analysis including only correct trials and only subjects with at least 20 trials per condition (n = 17) and a second analysis including all subjects and both correct and incorrect trials. Subsequent analyses investigating proficiency effects (according to the C-test and according to subjects´ error pattern in the GJT) were conducted without removal of incorrect trials because of the reasons mentioned earlier. Language proficiency effects (according to the C-test and according to subjects´ error pattern in the GJT) were also investigated.

Overall Analysis

Only correct trials



Anterior left ROI **Anterior midline ROI** **Anterior right ROI**

Figure 5.8 Grand Average waveforms of 17 subjects, shown on anterior midline, anterior right, and anterior left ROIs. Continuous line = correct condition, dashed line = incongruent condition, dotted line = pseudocongruent condition.

Visual inspection of the ERP patterns (cf. Figure 5.8) showed a substantial difference in mean area voltage for the pseudocongruent and incongruent condition in comparison with the correct condition in the P200 time window especially at anterior and midline electrode sites. The amplitude of the P200 was much more positive for the two incorrect conditions than for the correct condition.

The results of the statistical analysis are displayed in Table 5.18. Since the interaction of ROI with Pronoun Condition was almost significant, next, an analysis of the effects of Pronoun Condition within each ROI was carried out.

	df	*F*	*p*
ROI	5, 80	10.364	< .001*
Pronoun Condition	2, 32	7.890	.002*
ROI * Pronoun Condition	10, 160	2.330	.054

Table 5.18 Results of the GLM with the factors ROI and Pronoun Condition. Significant results are

		df	*F*	*p*
ROIs	**Posterior midline**	2, 32	7.412	.003*
	Posterior right	2, 32	1.006	.372
	Posterior left	2, 32	1.649	.210
	Anterior midline	2, 32	6.895	.004*
	Anterior right	2, 32	4.048	.032*
	Anterior left	2, 32	8.810	.001*

Table 5.19 The effects of Pronoun Condition within each ROI. Significant results are marked with an

marked with an asterisk. asterisk.

As can be seen in Table 5.19, Pronoun Condition had a significant effect in all anterior ROIs as well as the posterior midline ROI. Hence, in the next step, pairwise comparisons with *t*-tests were carried out only for these ROIs.

		df	*t*	*p*
Posterior midline ROI	**Correct – Pseudocongruent**	16	- 3.720	0.024*
	Correct – Incongruent	16	- 3.318	0.048*
	Pseudocongruent – Incongruent	16	.310	1.000
Anterior midline ROI	**Correct – Pseudocongruent**	16	- 2.737	0.180
	Correct – Incongruent	16	- 3.675	0.024*
	Pseudocongruent – Incongruent	16	.446	1.000
Anterior right ROI	**Correct – Pseudocongruent**	16	- 2.211	0.504
	Correct – Incongruent	16	- 2.709	0.180
	Pseudocongruent – Incongruent	16	- 0.192	1.000
Anterior left ROI	**Correct – Pseudocongruent**	16	- 3.499	0.036*
	Correct – Incongruent	16	- 3.525	0.036*
	Pseudocongruent – Incongruent	16	.500	1.000

Table 5.20 Results of pairwise comparisons with *t*-tests. Significant results are marked with an asterisk.

As displayed in Table 5.20, for the posterior midline ROI, *t*-tests showed that the difference between the correct and incongruent condition and between the correct and pseudocongruent condition was significant. For the anterior midline ROI, there was a significant difference in mean area voltage between the correct and incongruent condition. For the anterior right ROI, no significant differences were found. For the anterior left ROI, there was a significant difference in mean area voltage between the correct and incongruent condition as well as the correct and pseudocongruent condition.

Correct and incorrect trials

The results of the analysis were mostly similar to the previous analysis. The interaction of ROI with Pronoun Condition was significant ($F(10, 270) = 3.378$, $p = 0.007$) and the subsequent analysis showed that there was a significant main effect at the posterior midline ROI ($F(2, 54) = 11.718$, $p < 0.001$), anterior midline ROI ($F(2,54) = 10.599$, $p < 0.001$), anterior right ROI ($F(2, 54) = 10.705$, $p < 0.001$), and anterior left ROI ($F(2, 54) = 11.733$, $p < 0.001$). Results of the subsequent *t*-tests within these ROIs are displayed in Table 5.21.

This time, the difference in mean area voltage between the correct and incongruent condition and the correct and pseudocongruent condition was significant at all anterior sites and the posterior midline site.

		df	*t*	*p*
Posterior midline ROI	Correct – Pseudocongruent	27	- 5.503	< .001*
	Correct – Incongruent	27	- 3.671	.012*
	Pseudocongruent – Incongruent	27	- 0.022	1.000
Anterior midline ROI	Correct – Pseudocongruent	27	- 3.684	.012*
	Correct – Incongruent	27	- 3.935	.012*
	Pseudocongruent – Incongruent	27	.513	1.000
Anterior right ROI	Correct – Pseudocongruent	27	- 3.778	.012*
	Correct – Incongruent	27	- 3.822	.012*
	Pseudocongruent – Incongruent	27	.445	1.000
Anterior left ROI	Correct – Pseudocongruent	27	- 3.923	.012*
	Correct – Incongruent	27	- 3.969	< .001*
	Pseudocongruent – Incongruent	27	.759	1.000

Table 5.21 Results of pairwise comparisons with *t*-tests. Significant results are marked with an asterisk.

Summary

The results of the two analyses reported were similar. In both analyses there was a significant difference in mean area voltage between the correct and incongruent condition as well as the correct and pseudocongruent condition. This indicates a processing difference between correct and incorrect pronouns as early as 230 - 310 ms after pronoun onset.

Effects of language proficiency

Language proficiency according to the C-test

As before, in order to investigate effects of language proficiency, this analysis was conducted with subjects divided into a high- and low-proficient group according to their C-test scores. Upon visual inspection, mean area voltage in the P200 time window looked very similar for the two proficiency groups and indeed, the interaction of ROI with Condition and Level was not significant ($F(10, 260) = 0.944$, $p = 0.445$). Neither was there a main effect of Level ($F(1, 26) = 0.636$, $p = 0.432$). The results of the analysis with only correct trials were similar[17].

Language proficiency according to error pattern

Since there were no proficiency effects according to C-test scores, subjects were divided into two groups according to whether they showed L1 gender transfer in the error rates or not (cf. the analysis of the P600 component section 5.4.1, Figure 5.6). The two groups were named the “L1 gender transfer group” and the “no L1 gender transfer group”. Upon visual inspection, however, the P200

[17] Analysis only correct trials: The interaction of ROI with Condition and Level was not significant, ($F(10, 260) = 0.789$, $p = 0.571$), there was no main effect of Level $F(1, 26) = 0.047$, $p = 0.831$).

component for the two incorrect conditions seemed equally strong for the two groups, which was confirmed by the analysis. The interaction of ROI with Pronoun Condition and Group was not significant ($F(10, 260) = 0.612$, $p = 0.680$) and neither was there a main effect of Group ($F(1, 26) = 2.895$, $p = 0.101$). The results of the analysis with only correct trials were similar[18].

Summary

Contrary to the P600 analysis, there was no difference when subjects were divided into two groups according to their error pattern in the behavioral analysis. Regardless of whether they had showed L1 gender transfer in the error rates or sensitivity to grammatical violations in form of the P600 component, both groups exhibiteded an equally strong P200 in response to the incorrect pronouns.

Discussion P200

I observed a significant difference in the ERPs between the correct and incongruent condition and the correct and pseudocongruent condition in the P200 time window. The P200 had a greater, that is, more positive amplitude for the two incorrect conditions (the incongruent and pseudocongruent condition) than for the correct condition. This pattern corresponds to the P600 pattern discussed before. It is surprising that apparently even the low-proficient L2 learners recruited for the present experiment were sensitive to differences between correct and incorrect or, put differently, expected and unexpected anaphors as early as within the P200 time window.

The P200 is not a component that is typically looked for in language processing research investigating grammatical violations. It has mostly been reported in studies investigating perception and memory (Luck, 2005). In perception, for example, the P200 has been linked to the detection of certain stimulus features (Hillyard & Münte, 1984; Luck & Hillyard, 1994) and in memory to the recognition of old vs. new stimuli (Curran & Dien, 2003). In language processing, it has been shown to be sensitive to semantic expectancy effects (Federmeier & Kutas, 2002; Holcomb, Coffey, & Neville, 1992) and to the degree of semantic sentence constraint (Federmeier, Mai, & Kutas, 2005), usually in combination with an N400. These results taken together suggest that the P200 might indicate some kind of process that matches a perceived stimulus to (stored) mental representations (Evans & Federmeier, 2007), be it the features of a searched object, a previously encountered word, or an expectancy derived from sentence context. Furthermore, research has shown that lexical access already takes place as early as 200 ms after the presentation of a word (Dambacher, Kliegl, Hofmann, & Jacobs, 2006). Hence, even though the P200 does not belong to the most well-known components in the area of language processing, in the light of these results, it need not be surprising to find a P200 in the case of anaphor resolution. In the end, anaphor resolution is a task where the anaphor has to be correctly referenced to an antecedent stored in memory. However, in studies investigating expectancy and predictability effects, larger P200s have been obtained for expected stimuli and normal sentence completions than to unexpected stimuli and anomalous sentence completions (Federmeier & Kutas, 2002; Federmeier et al., 2005; Holcomb et al., 1992), while in the present experiment the P200 was larger for incongruent sentences.

Furthermore, besides the effects mentioned, the P200 has also been linked to frequency effects. Dambacher, Kliegl, Hofmann, and Jacobs (2006) studied frequency and predictability effects on ERPs during reading and found a more pronounced P200 in response to low-frequent words than to high-frequent words. They concluded that, since even the P200 as an early component was already

[18] Analysis only correct trials: The interaction of ROI with Pronoun Condition and Group was not significant ($F(10, 260) = 1.788$, $p = 0.112$). Neither was there a main effect of Level ($F(1, 26) = 2.065$, $p = 0.163$).

sensitive to frequency effects, lexical access to high-frequent words can happen very fast and at a very early processing stage. However, contrary to other studies, the P200 did not seem to be affected by contextual information or predictability (which is at odds with e.g., Sereno, Brewer, & O'Donnell, 2003). In addition, the finding that lower frequent words elicit a P200 with a more positive amplitude than high-frequent words is a common finding (Dambacher et al., 2006; Hauk & Pulvermüller, 2004; Rugg, 1990; Van Petten & Kutas, 1990). This offers another explanation for the P200 amplitudes found in the present experiment. The pronoun it used in the correct condition has a frequency class of 3[19] (Projekt Deutscher Wortschatz (PDW), cf. section 4.2.2). It is therefore somewhat more frequent in English than the pronouns *he* (class of frequency PDW = 4) and *she* (class of frequency PDW = 5), used in the incorrect conditions. Hence, another possible interpretation for the difference in P200 amplitude found between the pronouns in my experiment might be mere frequency effects.

The frequency explanation, contrary to the congruency explanation, seems even more plausible when looking at the results of the analysis of the "people filler sentences" (cf. section 5.2). For those fillers, sentences of the type *This is a mechanic/beautician. He/she/*it is friendly and patient.* involving stereotypically masculine (e.g., *mechanic*) as well as feminine (e.g., *beautician*) professions were used. Again, there were three conditions: the "stereotype congruent condition" with the anaphors *he* or *she*, respectively, the "stereotype incongruent condition" with the anaphors *she* or *he*, respectively, and the "grammatically incorrect condition" with the pronoun *it*. In the stereotype congruent condition, the pronoun was congruent with the stereotypically expected gender of the profession, while in the stereotype incongruent condition, it was not. The grammatically incorrect condition contained a grammatical violation. When one looks at the P200 amplitudes obtained for the different conditions (cf. Table 5.10), one can see that the grammatically incorrect condition (*it*) has the smallest amplitude, while the two grammatically correct conditions, the stereotypically congruent and incongruent conditions (*he* and *she*), have a bigger, more positive amplitude.

Note that in terms of correctness and congruency, this was exactly the opposite pattern from the one obtained in the presently discussed gender congruency experiment. There, I found a smaller amplitude for the correct condition and a more positive amplitude for the two incorrect conditions (pseudocongruent and incongruent condition). The picture starts to become somewhat clearer when one looks at the pattern not in terms of correctness or congruency but in terms of the influence of the different pronouns on the P200. Then, it becomes apparent that in both experiments the pronoun *it* elicits the smallest P200, independent of experimental condition. Note that the anaphor *it* was the critical word in the incorrect condition in the stereotype experiment but the critical word of the correct condition in the gender congruency experiment. The pronouns *he* and *she* are the critical words in the two correct conditions (congruent and incongruent) in the stereotype experiment but the critical words in the two incorrect conditions (pseudocongruent and incongruent) in the present experiment. Yet, *he* and *she* elicited a more positive P200 than *it* in both experiments. Taken together with the differences in frequency between the pronouns in English and the results by Dambacher et al. (2006), it seems likely that the P200 differences found are not due to congruency effects but are caused by frequency effects instead. Nevertheless, the experimental and people filler

[19] Frequency data were obtained from an German–English dictionary available under http://dict.uni-leipzig.de/ which forms part of the Projekt Deutscher Wortschatz (Quasthoff, Richter, & Biemann, 2006). A frequency class of 3, as given here, means that the English determiner the has got about 23 the number of occurrences than the selected word.

sentences involve anaphors referring to antecedents which differ in one important aspect, namely animacy. It cannot be completely ruled out that animacy has differential effects on P200 amplitude. Also, in the people filler sentences grammatical and world knowledge violations are mixed across conditions[20], which could lead to potentially differential results. Moreover, it is important to point out that the overall pattern found for the P200 resembles the pattern found for the P600 component discussed earlier, which revealed a difference in processing between grammatically correct and incorrect sentences.

Regarding proficiency effects, contrary to the P600 analysis, subjects showed an equally strong P200 in response to the incorrect and more infrequent pronouns *he* and *she* regardless of whether they had shown L1 gender transfer in the error rates or sensitivity to grammatical violations in form of the P600 component. Proficiency effects regarding the P200 have been reported by Weber-Fox, Davis, and Cuadrado (2003), who found a smaller P200 for high- than for normal-proficient (native English speaking) subjects in response to certain words, possibly reflecting retrieval processes and greater ease of processing. Nevertheless, in the present study, no such processing differences between lower and higher-proficient L2 learners were found. If the frequency explanation of the present P200 pattern is correct, it is a noteworthy finding that also very low-proficient L2 learners can be sensitive to frequency effects.

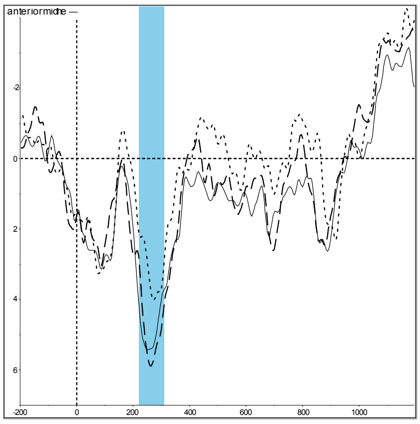


Figure 5.9 Grand Average waveforms of the "people filler sentences" (17 subjects and only correct trials), shown at the anterior midline ROI. Continuous line = congruent condition (*he/she*), dashed line = incongruent condition (*he/she*), dotted line = incorrect condition (*it*).

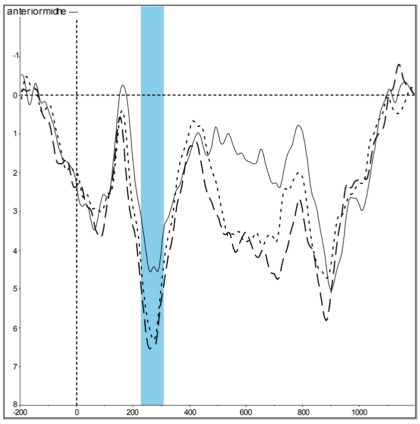


Figure 5.10 Grand Average waveforms of the experimental sentences (17 subjects and only correct trials), shown at the anterior midline ROI. Continuous line = correct condition (*it*), dashed line = incongruent condition (*he*/*she*), dotted line = pseudocongruent condition (*he*/*she*).

In conclusion, the interpretation of the P200 found in response to incorrect pronouns is not clear at first sight, because the P200 is not a component typically found in the processing of grammatical violations. Due to the occurrence of the P200 in response to violations of semantic expectancy and constraint, a congruency explanation seems plausible. Nevertheless, studies on semantic expectancy usually found the opposite pattern to the pattern found here, that is, a more pronounced P200 for expected stimuli than unexpected stimuli. Moreover, when considering the results of the filler sentence condition, the more pronounced P200 in response to the pronouns *he* and *she* favors a

[20] The filler sentence *This is a mechanic. *It is friendly and patient.* constitutes a grammatical violation, while *This is a mechanic. She is friendly and patient.* constitutes a violation of world knowledge or semantic expectancy.

frequency explanation. It is interesting that even the very low-proficient "L1 transfer group" was sensitive to frequency effects as shown by the P200.

5.4.3 Sustained negativity

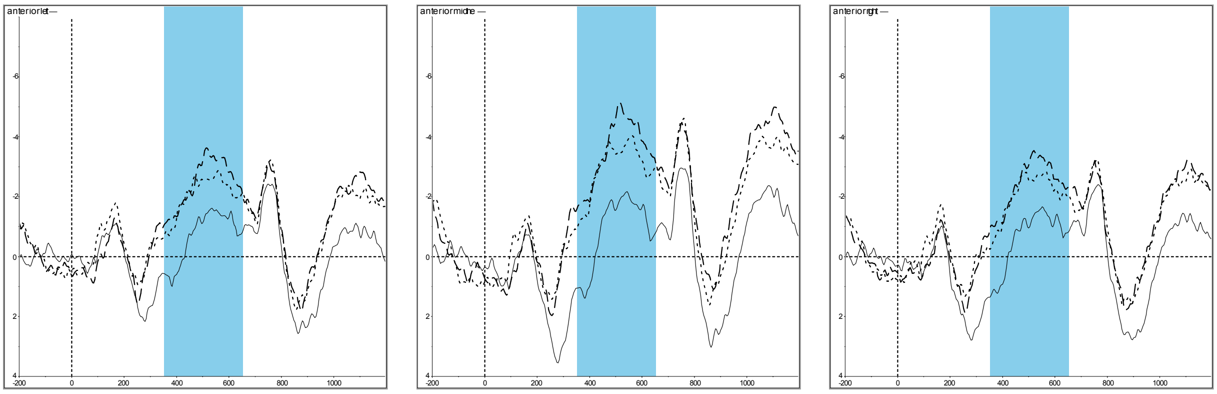
No predictions had been made regarding any components in response to the verb *is*. In fact, the occurrence of any components on stimuli after the pronoun stimuli containing the grammatical violations was unexpected. However, upon visual inspection of the whole sentence, a pronounced and prolonged negativity for the incorrect conditions was observed approximately in the time window of 350 - 650 ms, basically across the whole scalp. Since it is not clear which component was observed here, in the following, I will refer to this negativity as "sustained negativity" or "late negativity".

As before, the results of two analyses are reported, one analysis including only correct trials and only subjects with at least 20 trials per condition (n = 17) and a second analysis including all 28 subjects and both correct and incorrect trials. In the subsequent analyses, effects of language proficiency according to the C-test and then according to the pattern of error rates observed in the behavioral analysis were investigated. As before, these last two analyses were carried out with correct and incorrect trials included.

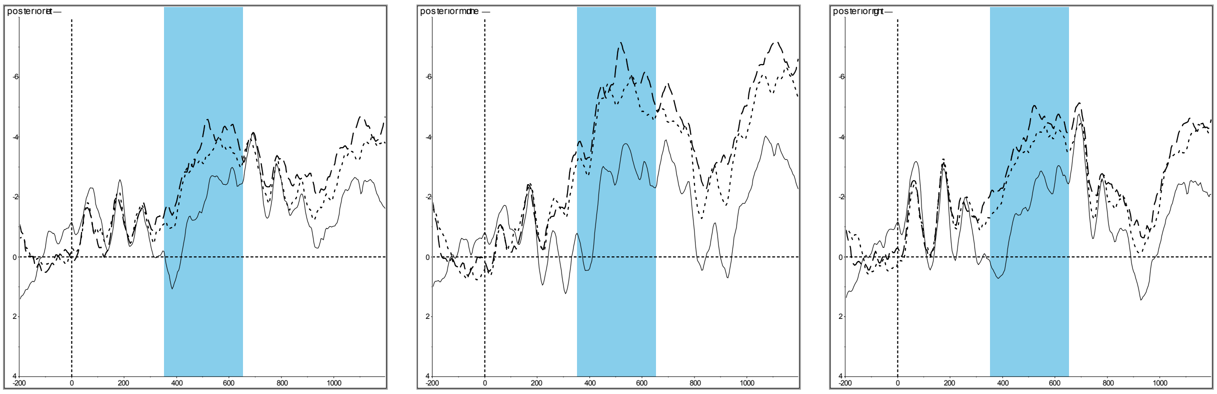
Overall analysis

Only correct trials

Visual inspection (Figure 5.11) showed that there was a greater negativity for the two incorrect conditions compared to the correct condition in a time window of approximately 350 - 650 ms after onset of the verb *is*. This negativity was present all over the scalp and seemed to be especially pronounced at midline sites.



Anterior left **Anterior midline** **Anterior right**



Posterior left **Posterior midline** **Posterior right**

Figure 5.11 Grand Average waveforms of 17 subjects and only correct trials, shown on anterior and posterior ROIs. Continuous line = correct condition, dashed line = incongruent condition, dotted line = pseudocongruent condition.

The results of the statistical analysis are displayed in Table 5.22. The interaction of ROI with Pronoun Condition was significant. Therefore, next, an analysis of the effects of Pronoun Condition within each ROI was carried out.

Table 5.23 shows that the effect of Pronoun Condition was significant at all posterior sites and at the anterior midline and anterior left ROI. In the following step, pairwise comparisons of Pronoun Condition were conducted with *t*-tests at each of these ROIs.

	df	*F*	*p*
ROI	5, 80	18.518	< .001*
Pronoun Condition	2, 32	7.654	.004*
ROI * Pronoun Condition	10, 160	3.500	.028*

Table 5.22 Results of the GLM with the factors ROI and Pronoun Condition. Significant results are marked with an asterisk.

		df	*F*	*p*
ROIs	**Posterior midline**	2, 32	9.096	.003*
	Posterior right	2, 32	8.937	.002*
	Posterior left	2, 32	6.991	.005*
	Anterior midline	2, 32	3.787	.047*
	Anterior right	2, 32	2.806	.090
	Anterior left	2, 32	3.804	.044*

Table 5.23 The effects of Pronoun Condition within each ROI. Significant results are marked with an asterisk.

As can be seen in Table 5.24, only the difference in mean area voltage between the correct and incongruent condition was significant, namely at the posterior midline and posterior right ROI. Contrary to my expectations after visual inspection of the data, the difference between the correct and pseudocongruent condition was not significant.

		df	*t*	*p*
Posterior midline ROI	**Correct – Pseudocongruent**	16	- 2.842	.180
	Correct – Incongruent	16	- 3.483	.045*
	Pseudocongruent – Incongruent	16	1.381	1.000
Posterior right ROI	**Correct – Pseudocongruent**	16	- 2.743	.210
	Correct – Incongruent	16	- 3.710	.030*
	Pseudocongruent – Incongruent	16	1.132	1.000
Posterior left ROI	**Correct – Pseudocongruent**	16	- 2.827	.180
	Correct – Incongruent	16	- 3.072	.105
	Pseudocongruent – Incongruent	16	.983	1.000
Anterior midline ROI	**Correct – Pseudocongruent**	16	- 1.515	1.000
	Correct – Incongruent	16	- 2.410	.420
	Pseudocongruent – Incongruent	16	1.512	1.000
Anterior left ROI	**Correct – Pseudocongruent**	16	- 1.568	1.000
	Correct – Incongruent	16	- 2.292	0.540

	Pseudocongruent – Incongruent	16	1.576	1.000

Table 5.24 Results of pairwise comparisons with *t*-tests (two-tailed, Bonferroni-corrected). Significant results are marked with an asterisk.

Correct and incorrect trials

The results were mostly similar to those of the previous analysis. I found a significant interaction of ROI with Pronoun Condition ($F(10, 270) = 3.748$, $p = 0.017$). In the next step, as shown in Table 5.25, there was a significant effect of Pronoun Condition at each ROI.

		df	***F***	***p***
ROIs	**Posterior midline**	2, 54	11.484	< .001*
	Posterior right	2, 54	8.521	.002*
	Posterior left	2, 54	8.273	.001*
	Anterior midline	2, 54	5.876	.008*
	Anterior right	2, 54	5.184	.013*
	Anterior left	2, 54	4.533	.018*

Table 5.25 The effects of Pronoun Condition within each ROI. Significant results are marked with an asterisk.

Consequently, *t*-tests comparing differences between conditions were conducted for each ROI. The results can be seen in Table 5.26

		df	***t***	***p***
Posterior midline ROI	**Correct – Pseudocongruent**	27	- 3.286	0.054
	Correct – Incongruent	27	- 3.853	0.018*
	Pseudocongruent – Incongruent	27	1.668	1.000
Posterior right ROI	**Correct – Pseudocongruent**	27	- 2.564	0.288
	Correct – Incongruent	27	- 3.472	0.036*
	Pseudocongruent – Incongruent	27	1.748	1.000
Posterior left ROI	**Correct – Pseudocongruent**	27	- 3.329	0.054
	Correct – Incongruent	27	- 3.363	0.036*
	Pseudocongruent – Incongruent	27	.959	1.000
Anterior midline ROI	**Correct – Pseudocongruent**	27	- 1.715	1.000
	Correct – Incongruent	27	- 3.009	0.108
	Pseudocongruent – Incongruent	27	2.112	0.792

Anterior right ROI	**Correct – Pseudocongruent**	27	- 2.202	0.648
	Correct – Incongruent	27	- 2.647	0.234
	Pseudocongruent – Incongruent	27	1.291	1.000
Anterior left ROI	**Correct – Pseudocongruent**	27	- 1.647	1.000
	Correct – Incongruent	27	- 2.626	0.252
	Pseudocongruent – Incongruent	27	1.633	1.000

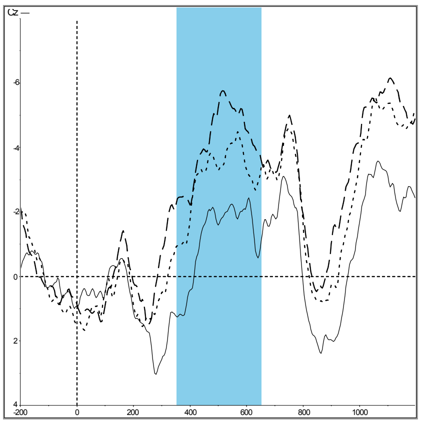
Table 5.26 Results of pairwise comparisons with *t*-tests (two-tailed, Bonferroni-corrected). Significant results are marked with an asterisk.

T-tests for the posterior midline and left region showed a significant difference between the correct and incongruent pronoun condition as well as an almost significant difference between the correct and pseudocongruent condition. There was also a significant difference in mean area voltages between the correct and incongruent condition in the posterior right region. When this analysis was carried out with only correct trials including all subjects, the results were similar but only the differences between the correct and incongruent condition were significant[21].

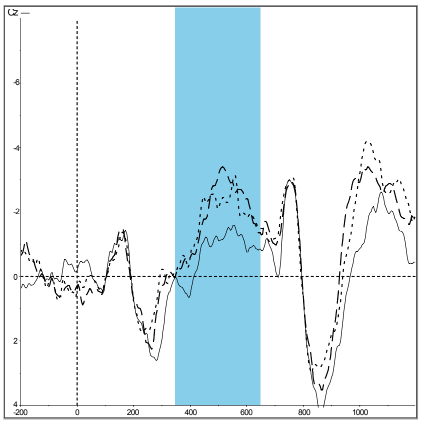
Effects of language proficiency

Language proficiency according to the C-test

Upon visual inspection (cf. Figure 5.12), it looked like there could be a processing difference between the two proficiency groups. For the high-proficient group (n = 14), the sustained negativity for the incorrect conditions seemed to be as pronounced as in the overall analysis, while for the low-proficient group (n = 14) the difference between the incorrect conditions and the correct condition looked very small. This observation was not confirmed by statistical analysis where no interaction effect of Level with ROI and Pronoun Condition was found ($F(10, 260) = 0.820$, $p = 0.474$). The result for the analysis with only correct trials was similar[22].



Cz electrode of the high-proficient group



Cz electrode of the low-proficient group

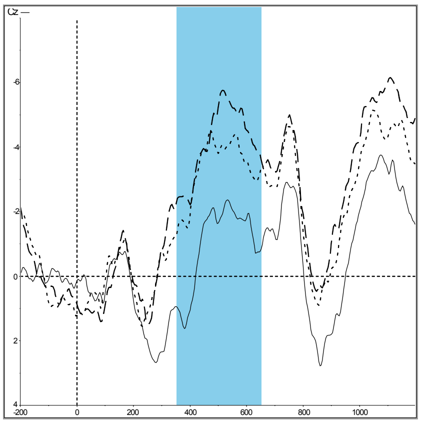
[21] Analysis only correct trials: Pairwise comparisons with *t*-tests carried out at posterior regions showed significant differences between the correct and pseudocongruent condition at posterior midline ($t(27) = -3.501$, $p = 0.018$), posterior right ($t(27) = -3.007$, $p = 0.054$) and posterior left ($t(27) = -3.287$, $p = 0.027$) ROIs.

[22] Analysis only correct trials: There was no interaction effect of Level with ROI and Pronoun Condition ($F(10, 260) = 1.503$, $p = 0.213$).

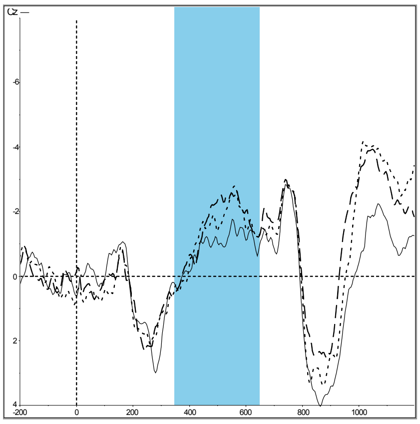
Figure 5.12 Grand Average waveforms of the high-proficient and low-proficient group, shown on the Cz electrode. Continuous line = correct condition, dashed line = incongruent condition, dotted line = pseudocongruent condition.

Language proficiency according to error pattern

Subjects were divided into two groups, according to whether they showed the expected error pattern (more errors in the pseudocongruent condition) or not. The “no L1 transfer group” (n = 15) did not show signs of L1 transfer, while the “L1 transfer group” (n = 13) experienced L1 gender transfer as shown in the behavioral data (cf. Table 5.5).



Cz electrode of the “no L1 transfer group”



Cz electrode of the “L1 transfer group”

Figure 5.13 Grand Average waveforms of the “no L1 transfer group” and the “L1 transfer group”, shown on the Cz electrode. The “no L1 transfer group” did not show the expected pattern in the error rates. The “L1 transfer group” showed the expected pattern in the error rates. Continuous line = correct condition, dashed line = incongruent condition, dotted line = pseudocongruent condition.

Upon visual inspection, the ERP patterns looked similar as in the previous proficiency analysis using C-test scores. The more proficient group, the “no L1 transfer group”, showed a strong negativity for the two incorrect conditions, while there was almost no greater negativity visible in the same time window for the less proficient “L1 transfer group”.

As can be seen in Table 5.27, statistical analysis confirmed that there was a significant interaction effect of Transfer Group with ROI and Pronoun Condition.

	df	*F*	*p*
ROI	5, 130	22.141	< .001*
ROI * Transfer Group	5, 130	1.752	.183
Pronoun Condition	2, 52	9.643	.001*
Pronoun Condition * Transfer Group	2, 52	2.056	.145
ROI * Pronoun Condition	10, 260	3.675	.016*
ROI * Pronoun Condition * Transfer Group	10, 260	3.626	.017*
Transfer Group	1, 26	4.832	.037*

Table 5.27 Results of the GLM with repeated measures factors ROI, Pronoun Condition, and Transfer Group ("no L1 transfer group" vs. "L1 transfer group"). Significant results are marked with an asterisk.

Subsequent analyses split up by Transfer Group showed that there was a significant interaction of ROI and Pronoun Condition only for the "no L1 transfer group" (cf. Table 5.28 and Table 5.29).

	df	*F*	*p*
ROI	5, 70	15.759	< .001*
Pronoun Condition	2, 28	8.021	.004*
ROI * Pronoun Condition	10, 140	5.245	.004*

Table 5.28 Results of the GLM for the "no L1 transfer group" with the factors ROI and Pronoun Condition. Significant results are marked with an asterisk.

	df	*F*	*p*
ROI	5, 60	7.569	.003*
Pronoun Condition	2, 24	2.361	.131
ROI * Pronoun Condition	10, 120	1.285	.295

Table 5.29 Results of the GLM for the "L1 transfer group" with the factors ROI and Pronoun Condition. Significant results are marked with an asterisk.

The analysis was continued for the "no L1 transfer group". A significant effect of Pronoun Condition was found for all posterior sites (posterior midline $F(2, 28) = 11.558$, $p = 0.001$, posterior right ($F(2, 28) = 11.375$, $p = 0.001$, posterior left $F(2, 28) = 11.558$, $p = 0.001$). For all anterior sites $p > 0.05$. Accordingly, *t*-tests comparing the different Pronoun Conditions were conducted for posterior sites only. Results are displayed in Table 5.30.

		df	t	p
Posterior midline ROI	Correct – Pseudocongruent	14	- 3.060	0.072
	Correct – Incongruent	14	- 4.005	0.009*
	Pseudocongruent – Incongruent	14	1.889	0.720
Posterior right ROI	Correct – Pseudocongruent	14	- 2.850	0.117
	Correct – Incongruent	14	- 4.412	0.009*
	Pseudocongruent – Incongruent	14	1.715	0.972
Posterior left ROI	Correct – Pseudocongruent	14	- 3.034	0.081
	Correct – Incongruent	14	- 3.686	0.018*
	Pseudocongruent – Incongruent	14	1.506	1.000

Table 5.30 Results of pairwise comparisons with *t*-tests (two-tailed, Bonferroni-corrected). Significant results are marked with an asterisk.

There was a significant difference between the correct and incongruent condition at all posterior ROIs. The difference between the correct and pseudocongruent condition was marginally significant at posterior midline and posterior left ROI. This result matches the results of the analysis with only correct trials[23].

Discussion

In the time window of 350 - 650 ms after onset of the verb *is*, significant differences between the correct condition and the two incorrect conditions were found at posterior ROIs. Contrary to the predictions (for the P600 component), no significant difference between the pseudocongruent and incongruent condition was found, but the observed pattern mirrors the pattern also found for the P600. When proficiency effects were investigated, no processing difference between proficiency groups divided by C-test scores were found. However, when subjects were divided into two groups according to whether they showed L1 transfer in their error rates, only the "no L1 transfer group" showed significant processing differences between the correct and incongruent condition and marginally significant differences between the correct and pseudocongruent condition. This showed that the "no L1 transfer group" processed the incorrect conditions differently from the correct condition. This is an almost native-like processing pattern with no evidence for L1 transfer. For the "L1 transfer group", on the other hand, no processing differences between the conditions could be observed. Hence, this group showed no sensitivity to L2 pronoun violations at all.

As mentioned above, under the hypotheses stated in section 5.1.1, no component had been predicted in response to the verb *is*. It is unclear why in addition to the processing differences observed at the time point of the pronoun violations further processing differences arose after

[23] Analysis only correct trials: *T*-tests carried out for the "no L1 transfer group" revealed a significant difference between the correct and incongruent condition at all posterior sites (posterior midline ($t(14) = -3.804$, $p = 0.018$), posterior right ($t(14) = -4.336$, $p = 0.009$) and posterior left ($t(14) = -3.513$, $p = 0.027$)). The difference between the correct and pseudocongruent condition was almost or marginally significant at all posterior sites (posterior midline ($t(14) = -3.231$, $p = 0.054$), posterior right ($t(14) = -2.957$, $p = 0.090$) and posterior left ($t(14) = -3.183$, $p = 0.063$)).

presentation of the verb *is.* If anything, it would seem more plausible to observe a component in response to the following adjective. Because only when the adjective appears, it was certain that the sentence was grammatically incorrect. After the introduction of the verb *is,* there are possible continuations that would be grammatical, such as *This is a bus. He/She is sitting by the window.* or *This is a bus. He/She is the bus driver.,* even if in these examples no referents are provided for the pronouns. Nevertheless, it seems that the possibility of finding a valid continuation for a sentence is not necessarily crucial for observing a component. It is possible that the posterior component found here in the time window of 350 - 650 ms after onset of the verb *is*, is simply a late component in response to the pronoun. Morgan-Short, Sanz, Steinhauer, and Ullman (2010), for example, found an N400 occurring in the usual time window in response to violations of grammatical gender agreement in low-proficient learners but also found "additional" N400-like effects in later time windows (600 - 900 ms and 900 - 1200 ms). These "late" N400 effects were considered continuations of the previous N400. According to the authors, such "elongated N400s" are not an uncommon finding in L2 speakers, who often show "long-lasting ERP effects" (p. 184, see also Hahne & Friederici, 2001; Hahne, Mueller, & Clahsen, 2006; Osterhout, McLaughlin, Pitkänen, Frenck-Mestre, & Molinaro, 2006). As a consequentce, it is possible that the negativity observed in the present experiment was an N400-like component in response to correct and incorrect pronouns, occurring in the time window of 950 - 1250 ms after pronoun onset.

Another interpretation is offered by Hammer, Jansma, Lamers, and Münte (2005), who also found an N400 (with a right-parietal maximum) time-locked to the word following a pronoun violation in a similar time window to the component observed here. Their study was conducted in German with native speakers. Interestingly, the late N400 was only found in response to the word following a pronoun violation when the antecedent involved an object (*Die Jacke$_{-fem}$ ist warm, weil sie$_{-fem}$/*er$_{-masc}$ gefüttert ist. = The jacket is warm, because she/he is lined*.), but not when the antecedent was a person (*Die Frau$_{-fem}$ ist beliebt, weil sie$_{-fem}$/*er$_{-masc}$ schön ist. = The woman was popular, because she/he is beautiful*.). According to the interpretation given by the authors, this means that semantic integration of persons was completed at the pronoun position, whereas semantic integration of objects continues beyond the pronoun. Therefore, in the case of a violation to "person-pronouns", the violation constitutes both a syntactic and a semantic violation and the process of pronoun resolution for these pronouns ends at the pronoun position. A violation of an "object-pronoun", on the other hand, constitutes a purely syntactic violation so that a semantic resolution could still be possible. Hence, the system continues to look for a way to integrate the pronoun. Since in the present experiment only inanimate objects were used as antecedents in experimental sentences, it is possible that the incorrect pronouns elicited such a late N400 component. Furthermore, the late N400 found by (Hammer et al., 2005) also mirrored the pattern of the previous P600 effect, just as in the present experiment.

Finally, a late negative component has also been found by other authors investigating monolingual and bilingual language processing. This negativity has been referred to as a "sustained negativity". Gillon Dowens, Vergara, Barber, and Carreiras (2010), for example, found a "sustained negativity" (p. 1877) with a broad distribution in monolingual and late L2 speakers which looks similar to the negativity found in the present experiment. The component was found in response to a number and a gender disagreement condition, compared to a correct baseline condition and occurred after a P600. It was observed at 1000 - 1300 ms after stimulus onset, which was similar to the time window in which the negativity was observed in the present experiment (950 - 1250 ms after pronoun onset). A similar-looking "sustained negativity" following a P600 has also been reported by Jiang, Tan, and

Zhou (2009), who investigated processing of violations involving the Chinese universal quantifier *dou* (all, every) in native speakers. They found a sustained negativity in a similar time window as in the present experiment, namely, from 300 to 800 ms post-onset of the following stimulus word in response to the violation (this experiment: 350 - 650 ms post-onset of *is*). The component was found on midline and lateral electrodes and was more pronounced in centro-parietal regions than in anterior regions, which is also similar to the distribution found in the present experiment. Based on their findings in subsequent experiments, they interpret this sustained negativity as an index of a second-pass process to reinterpret the sentence after a grammatical violation. A sustained negativity at frontal electrodes in the time window of 960 - 1500 ms has been found by Sabourin and Stowe (2008) in response to grammatical gender violations, in native speakers as well as one of the bilingual groups. The authors interpret the negativity similar to Sabourin and Stowe (2004), who conducted a study investigating memory effects in L1 sentence processing with ERPs. In this study, the observed frontal negativity was interpreted as reflecting working memory effort caused by the necessity to postpone sentence resolution due to an encountered grammatical violation. This interpretation would also be plausible in the case of the present experiment. However, the negativity found here had a more posterior distribution. In each case, it seems that even if the “sustained negativity” was not a well-established component up to now and the exact time window, distribution, and interpretation are not yet clear, it cannot be denied that this component is frequently found, especially in response to the stimulus following a violation (cf. also Coulson & Kutas, 2001; Otten & Van Berkum, 2009; Phillips, Kazanina, & Abada, 2005).

In conclusion, the negativity found in the present experiment is similar to the negativities found in other experiments. The present component seems very similar to the N400 found by Hammer et al. (2005), especially since it was also obtained in response to the stimulus following a pronoun violation and mirrored the pattern of the previous P600. Furthermore, the present component also has some similarities with a “sustained negativity” found by other authors (Gillon Dowens et al., 2010; Jiang et al., 2009; Sabourin & Stowe, 2004, 2008). According to the interpretations offered in the literature, the late negativity observed here could reflect an ongoing attempt to successfully resolve the pronoun violation (Hammer et al., 2005; Jiang et al., 2009) or the working memory efforts associated with this (Sabourin & Stowe, 2008). The incorrect conditions are processed differently from the correct conditions and when subjects are divided into two groups, the “L1 transfer group” and the “no L1 transfer group”, this processing difference was only found in the more proficient group, that is, the “no L1 transfer group”. The “L1 transfer group”, on the other hand, showed no sensitivity to L2 grammatical violations. This was identical to the findings in the P600 analysis. Hence, it seems that the negativity found here can be attributed to the gender violations at the pronoun position and that the integration process started at the pronoun position, as indicated by the P600, is continued.

5.5 Discussion Experiment 2

The present experiment investigated L1 gender transfer effects in a sentence processing task, measuring error rates and ERPs. ERPs can give information on the time course of information integration and parsing during language processing. Experiment 1 (cf. chapter 4) showed that, different from the research focus in the present literature, it was not only L1 characteristics that impact transfer processes but that also L2 characteristics, such as gender transparency, play a role. Therefore, the presence or absence of certain L1 features in the L2 might be important. The central question in this second experiment was whether L1 gender transfer would also be possible into an L2 lacking gender, namely, English. Subjects were native speakers of German who were very low-proficient in their L2 English.

It was hypothesized that gender transfer effects would become apparent in the error rates. The highest error rates were predicted for the pseudocongruent condition, lower error rates were predicted for the correct condition, and the lowest error rates were predicted for the incongruent condition. Regarding ERP components, the focus was on the P600 component as an indicator of the detection of syntactic and morphosyntactic violations (cf. sections 1.3.2 and 1.5.1). The weakest P600 was expected for the correct condition, as this condition was correct in the L2 and should therefore not give rise to any syntactic re-analysis processes. The P600 in response to the pseudocongruent condition was expected to lie in between the correct and incongruent condition, as it was hypothesized to be perceived as somewhat correct due to L1 gender transfer. If gender transfer occurs, the pseudocongruent condition should be processed significantly different from the incongruent condition. If there was no gender transfer, the pseudocongruent condition should be processed similarly to the incongruent condition.

One of the most important findings in the present experiment was that clear L1 transfer effects were obtained in the error rates. As predicted, significantly more grammatical judgment errors were made in the pseudocongruent condition than in the two other conditions. This showed that L1 gender transfer can even take place when the L2 lacks gender. Furthermore, as hypothesized, less errors were observed in the incongruent condition than in the correct condition. This was probably due to the fact that only the incongruent condition can be thought of as free of L1 gender transfer, as only in this condition the second sentence would have been incorrect in both L1 and L2. Surprisingly, however, despite the pattern in the error rates providing such strong evidence for L1 gender transfer, these transfer effects in the error rates were not observed in the ERP patterns. On the contrary, in the P600 component, processing differences were observed between the correct and the two incorrect conditions. No difference between the pseudocongruent and incongruent condition was observed. This effect would have been expected in the case of absence of L1 gender transfer or for a native control group. The P600 seemed “native-like” in distribution, latency and amplitude. Nevertheless, it is difficult to estimate if this was really a completely “native-like” P600 as it could not be directly compared to a native control group. Furthermore, as predicted for these low-proficient speakers, no LAN was obtained. Since in early stages of language learning L2 speakers sometimes exhibited an N400 instead of a P600, it was speculated that also an N400 could be observed. However, this was not the case.

As a result, gender transfer was manifested in the error rates but did not become manifest in the ERPs. To shed light on these seemingly contradictory results and in order to investigate proficiency effects, subjects were divided into two groups according to whether they had shown L1 transfer in the error rates or not: the “L1 transfer group” and the “no L1 transfer group”. It was thought that the “L1 transfer group” would potentially exhibit the hypothesized pattern also in the ERP results. However, surprisingly, the lower-proficient “L1 transfer group” showed no sensitivity to L2 grammatical violations at all, as no P600 became evident for any of the conditions. The “no L1 transfer group”, on the other hand, showed the same pattern that appeared in the overall analysis, that is, an equally pronounced P600 for the two incorrect conditions. Besides this, no other effects of proficiency were observed. When subjects were divided into two proficiency groups according to their C-test scores, no differences between the two groups were found.

Furthermore, the P600 processing pattern was mirrored in two other components, which are not typically reported for grammatical violations, namely, the P200 and a (sustained) negativity component in response to the verb *is*, following the critical pronoun. The P200 does not belong to the well-established components in the processing of semantic or syntactic violations, like the N400, P600, LAN, and ELAN. However, in studies investigating perception and memory the P200 seemed to indicate some kind of matching process that compares features of the encountered stimulus with the features of a target stimulus stored in memory. Therefore, it seems reasonable to assume that it could also be sensitive to the matching of an anaphor with its antecedent. However, in studies investigating expectancy and predictability effects the amplitude pattern for the P200 was opposite to the pattern found for the present conditions (Federmeier & Kutas, 2002; Federmeier et al., 2005; Holcomb et al., 1992). Another possibility is that the more pronounced P200 in response to the incorrect pronouns reflects frequency effects, as stronger P200s are also found for low-frequent words than for high-frequent words (Dambacher et al., 2006; Hauk & Pulvermüller, 2004; Rugg, 1990; Van Petten & Kutas, 1990) and the pronouns *he* and *she* are less frequent in English than the pronoun *it*. The comparison with ERP patterns of the people filler pronouns renders this interpretation more likely. Yet another possibility is that a combination of congruency and frequency effects is at work here. Regarding proficiency effects, it is important to point out that for the P200 component, the “L1 transfer” and “no L1 transfer group” showed the same processing pattern, that is, the overall pattern found. This demonstrates that even low-proficient subjects, who exhibited L1 transfer in the error rates and no sensitivity to L2 violations in the P600 time window, are possibly sensitive to frequency effects.

Also the occurrence of a negativity in response to the verb *is,* following the critical pronoun, was not completely clear. The negativity occurred in the time window of 350 - 650 ms after the verb *is*, which is equal to 950 - 1250 ms after pronoun onset, in response to the two incorrect conditions, compared to the correct condition. In principle, it could be a late manifestation of an N400 component, which appears to be not unusual in the L2 processing literature (Morgan-Short et al., 2010). This seems especially plausible considering the results of Hammer, Jansma, Lamers, and Münte (2005). These authors found a late N400 in response to pronoun violations when the antecedent was a thing, as it was also the case in the present experiment. Furthermore, prior to this late N400, also a P600 had emerged, just as in the present experiment. Taken together with the interpretation of other late or “sustained” negativities(Gillon Dowens et al., 2010; Jiang et al., 2009; Sabourin & Stowe, 2004, 2008), this negativity could reflect an increased attempt to somehow integrate the encountered incorrect stimulus word or the working memory load associated with it. Most importantly for the present discussion, this late negativity basically reflects the results found for the P600: The incorrect conditions are processed differently from the correct condition, which was observed in the overall analysis as well as in the analysis of the “no L1 transfer group”. Only the lower-proficient “L1 transfer group” showed no difference between the correct and incorrect conditions, just as in the P600 analysis.

Overall, regarding the ERP results, the predictions were not borne out. The pseudocongruent condition was not processed differently from the incongruent condition in any of the observed components. On the contrary, the pseudocongruent condition was processed very similarly to the incongruent condition, but significantly different from the correct condition. As mentioned before, this is the processing pattern that is expected in the absence of L1 gender transfer. Yet, as shown in the error rates, L1 gender transfer did occur. The error patterns were used as a basis for creating two groups, the “L1 transfer group” and the “no L1 transfer group” in order to clarify why the gender transfer in the error rates did not become apparent in the ERPs and with the addition goal of investigating proficiency effects. In the ERPs, the more proficient “no L1 transfer group” showed a native-like processing pattern across all three components. The “L1 transfer group”, on the other hand, showed no sensitivity to L2 grammatical violations regarding the P600 and late negative component. Here, all conditions were processed in the same way. It seems that some of my low-proficient participants had already developed a sensitivity for this kind of syntactic violations, whereas others still had not. My data suggest that low-proficient speakers go through different phases in the language acquisition process. At least in some cases, L1 gender transfer can be minimized and native-like sensitivity for L2 gender violations can be developed even at a low proficiency level. Regarding the P200 component, however, the same processing differences were observed in the “L1 transfer group” and in the more proficient “no L1 transfer group”, that is, the incorrect conditions (including the less frequent pronouns) were processed differently from the correct condition. This probably shows that even low-proficient speakers exhibit sensitivity to word frequency. Furthermore, when proficiency effects were investigated using C-test scores, no effects were found, neither in the behavioral, nor in the ERP data. So it seemed that using behavioral data as a basis for creating proficiency groups can be a fruitful approach (cf. also Osterhout, McLaughlin, Pitkänen, Frenck-Mestre, & Molinaro, 2006).

Nevertheless, it remains puzzling why no evidence for gender transfer became apparent in the ERPs of the “L1 transfer group” as an online processing measure but only in the behavioral accuracy measure. This seems especially strange because, evidently, the behavioral output is a result of the preceding processing effort. In addition, as stated before, ERPs are thought to be more sensitive than behavioral measures (Luck, 2005; McLaughlin et al., 2004; Tokowicz & MacWhinney, 2005) as they give insights into processing as it unfolds in the brain. Behavioral measures, on the other hand, can only give information on the “end product” of such processing so that what happens prior to the output remains a “black box”. Therefore, since behavioral output was a result of online processing which was measured by ERPs, transfer effects that became apparent in the error rates should also have manifested themselves in the ERPs.

Yet, by and large, the present results are in line with the results of McLaughlin, Tanner, Frenck-Mestre, Valentine, and Osterhout (2010). They found that L2 speakers´ behavioral sensitivity correlated with the P600 amplitude, that is, L2 learners with stronger behavioral sensitivity showed a more robust P600 than those with less behavioral sensitivity, who showed only a small or no P600 (p. 126). Aside from L1 transfer, this is more or less in agreement with the present results: Lower error rates (“no L1 transfer group”) also correlated with a more pronounced P600, while the “L1 transfer group”, exhibiting higher error rates, showed no P600. However, contrary to McLaughlin et al. (2004), in the present experiment no intermediate stage between no L2 sensitivity and grammaticalization was found. Furthermore, the present results are also in line with the findings of Foucart and Frenck-Mestre (2011, experiment 1), who found that only a part of the native German speakers was sensitive to gender violations in L2 French when nouns had different gender values across languages, while all L2 learners were sensitive to gender agreement violations in the case of gender-congruent nouns (p. 387). In the present experiment, all nouns were gender-incongruent. In addition, Foucart and Frenck-Mestre (2011) concluded that even L2 learners who appear relatively homogeneous in proficiency and other factors might show different learning rates for a specific grammatical structure (p. 388). This is consistent with the results by McLaughlin et al. who state that “We also show that although learners’ brain responses are quite variable, this variability is highly systematic and can be used to identify meaningful subgroups of learners.” (p. 124). The fact that individual differences can arise even in groups of homogeneous proficiency is less surprising when one considers that ERP patterns can rapidly change for various reasons, potentially yielding individual differences. For example, as discussed in section 1.4.2, changes in ERPs are possible after only short periods of training, even regarding such difficult structures as grammatical gender (Davidson & Indefrey, 2009). Furthermore, it has also been found that ERP patterns can suddenly become native-like after a certain time of non-exposure (Morgan-Short, Finger, Grey, & Ullman, 2012). In addition, subtle factors such as the type of instruction (implicit vs. explicit) can have an influence on the native-likeness of ERP patterns (Morgan-Short et al., 2010). In view of all these factors that can affect ERP patterns in L2 processing, it becomes clear that most groups of L2 learners will probably display individual differences in ERPs, no matter how homogeneous they seem to be according to collected metadata.

What do the present results mean in light of the literature discussed in the introduction of the present experiment? Different from Barto-Sisamout, Nicol, Witzel, and Witzel (2009), in the present experiment clear evidence was provided that L1 influences even arise in cases where an L1 morphosyntactic feature is absent in L2. Barto-Sisamout et al. had only obtained a trend towards slower reading times in a comparable condition, that is, a "condition in which morphological marking is required in the L1 but not in the L2" (p. 1) but no significant result. Similar to the results of the pronoun production task conducted by Antón-Méndez (2010), it was shown that L2 speakers can have difficulties with L2 pronoun processing and that L2 pronoun processing was biased by L1. Consistent with the results of the eye-tracking task by Conklin, Dijkstra, and van Heuven (2007), I also found evidence for gender transfer effects from a gendered language into an ungendered language in a pronoun resolution context. However, contrary to the earlier discussed studies providing evidence for L1 influences in gender processing in ERPs (Foucart & Frenck-Mestre, 2011; Frenck-Mestre et al., 2009) and evidence for L1 gender transfer in L2 processing (Ganushchak et al., 2011; Midgley et al., 2007), the present experiment failed to show evidence of L1 gender transfer in an ERP component. Nonetheless, the situation in the experiments of Foucart and Frenck-Mestre (2011, experiments 2 and 3) and Frenck-Mestre et al., (2009) might have been a little different: Only general L1 influences in L2 gender processing instead of congruency effects were observed, as native German speakers proved to be insensitive to L2 gender violations only when these involved plural forms – probably because German lacks gender agreement for plural but not singular forms. Furthermore, similar to the present result, in the study by Ganushchak et al. (2011) also a transfer effect in the error rates was obtained. But here, subjects were in a language-mixing context and put in response conflict, which probably favored the rise of interference effects. In addition, a different ERP component than in the present experiment was investigated. The stimuli used by Midgley et al. (2007), on the other hand, probably most closely resembled the stimuli used in the present experiment, as also pronouns were used as critical items. However, since participants read for comprehension, error rates did not give insights on possible gender interference processes. A difference between conditions in the P600 time window was found, but the P600 had an unusual anterior distribution. Hence, the results of the present experiment only partly fit into the literature discussed in the introduction. In light of the literature, it is uncertain why the present experiment showed L1 transfer effects in the error rates but not in the ERPs. In each case, one possible explanation, namely, that the participants in the present experiment might have been too high-proficient for L1 transfer effects to arise can be ruled out, since regarding the ERP results, two proficiency groups could be identified: a lower-proficient group showing no sensitivity to L2 gender violations and a higher-proficient group showing a native-like processing pattern. If there was an intermediate stage using the L1 as a processing basis, it should have become apparent in the ERP data.

Regarding the models discussed in section 1.5, the Declarative/Procedural Model (DP Model) and the Competition Model, only some of their predictions are supported. As explained in section 1.5.1, according to the DP Model, contrary to native speakers, late L2 learners tend to rely on declarative structures instead of procedural structures when processing L2 grammar. The N400 has been hypothesized to indicate the usage of declarative structures, while the LAN has been hypothesized to indicate more automatized processing and the usage of procedural structures (Morgan-Short et al., 2010), depending on left-frontal structures. As mentioned in section 1.3.2, the P600 seems to indicate more controlled processing and structural reanalysis. In the present experiment, no N400 indicating reliance on declarative structures has been obtained, different from experiments showing that in the beginning, learners might tend to rely on declarative structures instead of procedural structures (Guo et al., 2009; McLaughlin et al., 2010; Morgan-Short, Steinhauer, Sanz, & Ullman, 2012; Osterhout, McLaughlin, Pitkänen, Frenck-Mestre, & Molinaro, 2006; Steinhauer, White, & Drury, 2009; cf. sections 1.3.2 and 1.4.2. Furthermore, no LAN, which is sometimes observed in response to morphosyntactic violations in native speakers and indicates clear reliance on procedural structures, has been observed, either. Since no native control group has been tested, it is not clear whether a LAN would have been observed in native processing. The (possibly) native-like P600 found in the present experiment indicates reliance on L1 neurocognitive processing systems (Morgan-Short et al., 2010) but contrary to the LAN, it does not necessarily indicate reliance on procedural memory (Newman, Ullman, Pancheva, Waligura, & Neville, 2007; Ullman, 2004). Other, previously cited behavioral studies investigating gender agreement had found that late L2 learners were not able to rely on procedural structures (Blom, Polisenska, & Weerman, 2008; Kempe, Brooks, & Kharkhurin, 2010; Sabourin, Stowe, & de Haan, 2006; cf. section 1.5.1). However, in the case of the present data, the conclusion regarding the usage of the declarative vs. procedural memory system remains uncertain. Moreover, the studies conducted by Blom et al. (2008) and Sabourin et al. (2006) are in line with the earlier discussed finding that L2 performance deteriorates with greater agreement distance (cf. section 2.2). Also, Morgan-Short et al., (2010) found a more reliable N400 for local dependencies than for non-local dependencies, which was in line with the DP model stating that local dependencies should be easier to learn in declarative memory than non-local dependencies (p. 182). In the present study, however, a non-local dependency was tested and a possible native-like P600 was found. Hence, the results of the present experiment suggest that L2 performance does not necessarily have to be worse than native-like at greater agreement distances, even at beginning stages. At least for one subgroup of the participants, performance in behavioral and ERP measures was (almost) native-like. Hence, it might be the case that, when the L2 structure is sufficiently simple, native-like processing is possible, even at low proficiency levels, for L2 grammatical gender and at greater agreement distances.

As stated in the introduction of the present experiment, the Competition Model would predict the most pronounced P600 in the case where a structure is grammatically incorrect in both L1 and L2, as consequence of cue summation (Hernandez, Li, & MacWhinney, 2005; p. 4). Because of this, for the present experiment it was predicted that the P600 would be less pronounced for the pseudocongruent condition than for the incongruent condition, as in the pseudocongruent condition L1 cues would contradict L2 cues, which was not borne out. The ERP results of Tokowicz and MacWhinney (2005) (cf. section 1.7.2 on the Competition Model), however, supported the predictions made by the Competition Model. Native English speakers who were low-proficient L2 learners of Spanish exhibited a P600 to grammatical violations that were similar to L1 or unique to L2 but not to structures that differed between L1 and L2. Interestingly, and also different from the present results, grammatical judgments were at chance for all constructions. So at first sight, the present result which revealed L1 gender transfer in the error rates (especially for one subgroup of participants) but not in the ERP data seemed to be at odds with the findings of Tokowicz and MacWhinney. Nevertheless, it is possible that the structure tested in the present experiment can be considered a structure that is “different” in the L2. Then the result of the “L1 transfer group” would be roughly comparable to the result found in the “different condition” by Tokowicz and MacWhinney: high error rates in the behavioral task and no sensitivity to grammatical violations in the ERPs. Furthermore, the results of the present experiment are in line with the general predictions of the Competition Model, which states that all languages of a speaker will always compete, especially at lower proficiency levels. This was the case for the “L1 transfer group”. Furthermore, competition is supposed to be overcome at higher proficiency levels due to greater L2 resonance and entrenchment, as was the case for the higher-proficient “no L1 transfer group”.

In conclusion, the results of the present experiment show that L1 gender transfer is possible even into a language with an extremely simple gender system. These transfer effects are clearly mediated by proficiency: Transfer effects were present at low proficiency levels and were overcome at higher proficiency levels. Therefore, the results further showed that native-like processing is possible even for L2 gender and at greater agreement distances (if the L2 gender system was sufficiently simple). These results point towards a shared bilingual gender system, at least in low-proficient speakers, and provide evidence for interference in L2 online processing. More generally speaking, regarding language transfer, it was shown that transfer of L1 features that are not present in L2 is nevertheless possible.

6. Overall discussion

The present thesis attempted to investigate the circumstances under which L1 gender transfer occurs in late bilinguals and which factors mediate this transfer. One aim of the present thesis was to describe how gender transfer changes in the course of language learning and how it changes with increasing proficiency. Another question to be answered was how L1 transfer is mediated by L2 factors, such as the transparency and complexity of the L2 gender system.

These questions were investigated empirically with two experiments. In the following, I will first briefly summarize the setup and the most important results of the two experiments conducted. I will then explain what these experiments revealed regarding the central questions asked in the present thesis. In the end, I will discuss the interpretation of my results regarding open questions from the research literature. I will also describe the implications for the two models of bilingual language processing introduced in section 1.5.

With the aim of investigating L1 transfer effects in L2 gender acquisition, two experiments with different language pairs manipulating several factors such as task demands, online/offline mode, agreement distance, cognate status, and language proficiency were conducted. The first experiment consisted of an LDT and PNT and an offline gender assignment task. Previous studies had shown that gender interference effects can be observed with these types of paradigms. The LDT and PNT were conducted in order to investigate online transfer processes. In combination with the offline task, it should be possible to estimate the amount of L1 influences stemming from faulty representations (acquisition-based account, cf. section 3.3.2 and Lemhöfer, Schriefers, & Hanique, 2010 and Lemhöfer, Spalek, and Schriefers, 2008) rather than online interference processes (online-account). Another goal of using these different methodological approaches was to investigate the influence of task demands. The syntactic structures in question were NPs and bare nouns. Target languages were L2 Spanish and L2 German. Since for Germanic languages so far the monolingual gender interference effect has only been observed in NPs (Schriefers & Teruel, 2000; Schriefers, 1993; cf. section 3.3.2) and for Romance languages only in bare nouns (Cubelli et al., 2005; Paolieri et al., 2011; Paolieri, Lotto, et al., 2010), interference effects were investigated for both syntactic structures. Spanish and German were chosen for several reasons: First of all, so far gender interference effects have only been investigated within Romance languages (Morales et al., 2011; Paolieri, Cubelli, et al., 2010) or within Germanic languages (Lemhöfer et al. 2010; 2008), but not across these two language families (cf. section 3.3.2). It has already been shown that language transfer is possible across language families with symmetric gender systems (Bordag & Pechmann, 2007, L1 Czech–L2 German; Salamoura & Williams, 2007b, L1 German–L2 Greek) and within language families with asymmetric gender systems (Lemhöfer et al. 2010; 2008; both L1 German–L2 Dutch). But it is not known whether transfer is also possible between languages from different language families with an asymmetric gender system. Furthermore, the Spanish and German gender systems differ regarding transparency and complexity. Spanish has a two-way and fairly transparent gender system, while the German three-way gender system is much more intransparent. German gender is generally difficult to acquire, also because of the complex declensional system which obscures form–meaning mappings regarding grammatical gender for the L2 learner (Rogers, 1987; Taraban, McDonald, & MacWhinney, 1989; cf. section 4.1.1). Native speakers of Spanish and German who spoke German and Spanish as an L2, respectively, took part in the experiment. Tasks were conducted in the L2 and low-proficient as well as high-proficient subjects were tested. Two native control groups also participated.

Noncognates and cognates were used as stimuli because in previous studies interference effects have been observed to be stronger for cognates than for noncognates (Lemhöfer et al., 2010, 2008; Salamoura & Williams, 2007, only error rates). RTs as well as error rates were measured.

Experiment 1 showed that gender transfer between a Germanic and Romance language with asymmetric gender systems is possible. This became apparent especially in the error rates of the offline gender assignment task for the Spanish bilinguals and in the error rates of the PNT for both bilingual groups[1]. Importantly, in the offline gender assignment task, only transfer effects for Spanish subjects, or put differently, for L2 German were observed. For the German subjects, on the other hand, no transfer effects into L2 Spanish were found. This is probably due to the greater transparency of the Spanish gender system. Follow-up analyses of the error rates of the German subject group showed that error rates were greatly influenced by gender transparency. Significantly more gender assignment errors were committed for intransparent than for transparent Spanish nouns. In addition, error rates were low even for unknown transparent nouns which shows that gender transfer is strongly affected by transparency. In the offline gender assignment task, robust transfer effects were observed for the Spanish subjects into L2 German. This suggests that gender transfer is stronger in the case of an L2 with an intransparent gender system. In addition, in the error rates of the PNT, interference effects were stronger for Spanish subjects than for German subjects and Spanish subjects also committed many determiner errors in NP naming. The longer RTs for NP naming than bare noun naming further showed that Spanish subjects had difficulties retrieving the correct article in German. German subjects, on the other hand, made few determiner errors and RTs between naming conditions (bare noun vs. NP naming) were similar. This shows that they had few problems retrieving the correct Spanish article, even in an online production task.

Regarding proficiency effects, low-proficient subjects made more gender errors than high-proficient subjects in the PNT as well as the offline gender assignment task. However, little effect of proficiency on gender transfer was found. In the offline gender assignment task, proficiency did not seem to mediate gender transfer effects. This might be due to the Germans performing almost at ceiling in both proficiency levels, while the Spanish subjects experienced difficulties across both proficiency levels. In the error rates of the PNT in Spanish, proficiency might have had some influence on gender transfer. Transfer effects seemed to be present in the low-proficient group but not in the high-proficient group. However, because of the very low overall error rate especially in the high-proficient group, this result is probably not very robust. Furthermore, due to the failure to find clear interference effects in the RTs of the online tasks, the effects of different task demands in gender transfer could not be determined.

The second experiment investigated gender transfer in sentence processing involving pronoun resolution. ERPs and error rates were analyzed. Subjects´ L1 was German and their L2 English. In Experiment 1 it was shown that L2 factors, such as gender transparency, are important. Gender interference was greatly reduced in transparent L2 nouns. Therefore, Experiment 2 tried to establish whether gender transfer occurs at all into a language that lacks grammatical gender and could therefore be considered an “extremely transparent” gender system. This time, a sentence processing task involving gender agreement over a greater agreement distance, namely, in pronoun resolution, was conducted. Subjects were low-proficient in English. Furthermore, in addition to error rates, ERPs were employed as the dependent measure. ERPs, as an online measure, are thought to be more

[1] Even though these effects were only significant in the analyses by participants.

sensitive than behavioral measures and can give information on the time-course of language processing as it unfolds in the brain.

The second experiment showed that grammatical gender transfer into a language without a grammatical gender system is possible. This confirms findings from the literature and the prediction made by the Competition Model that the L1 is always activated and influences L2 processing. Apparently, the Competition Model dictum “whatever can transfer will” (MacWhinney, 2005a, p. 55) can be applied to grammatical gender: “gender is transferred, no matter what”. Furthermore, regarding transfer in general, the second experiment showed that transfer of an L1 feature that is not present in the L2 is nevertheless possible. This became apparent in the error rates but not in the ERPs. ERPs, on the other hand, revealed that one group of the low-proficient subjects, who had shown no L1 transfer effect in the error rates, already exhibited a native-like processing pattern in terms of the P600. Hence, this group displayed no L1 transfer but rather differential processing patterns regarding grammatical and ungrammatical sentences. Another subject group, who had shown strong L1 transfer effects in the error rates, however, also showed no evidence for L1 transfer in the ERPs, either. Instead, this apparently lower-proficient group exhibited no evidence for sensitivity to L2 violations at all. This suggests that L2 learners go through different stages in L2 learning. At least in some cases L1 transfer can be minimized already at an early stage and sensitivity for L2 gender violations can be developed. It is interesting that even in this seemingly homogeneous group of low-proficient L2 learners, different subgroups who are in different acquisition stages regarding the grammatical structure in question can be identified.

Furthermore, the present results show that L2 learners can process L2 gender agreement successfully at greater syntactic distances and overcome L1 transfer. This is contrary to the statement by Blom, Polisenska, and Weerman (2008; cf. sections 1.7.1 and 4.12) who maintained that late L2 learners only disposed of lexical strategies to process L2 grammatical gender which would prove deficient at greater agreement distances. Apparently, at least when the L2 gender system is simple enough, processing across greater agreement distances can be native-like.

As stated before, research on language transfer had so far concentrated more on the influence of L1 characteristics (cf. section 5.1). The first experiment of the present thesis showed that also L2 characteristics such as noun ending transparency play a role. In the case of the transparent Spanish gender system, no transfer became evident even at lower proficiency levels. In the case of the more intransparent German gender system, on the other hand, gender transfer was apparent at lower and higher proficiency levels. The second experiment demonstrated that transfer can still occur when the L2 gender system is very simple or very transparent, at least at greater agreement distance and low language proficiency. However, the second experiment also indicates that in this case, L1 gender transfer can be quickly overcome at only slightly higher proficiency levels.

The present thesis aimed to describe the circumstances under which L1 gender transfer occurs. Across the two experiments reported, different factors thought to possibly influence L1 gender transfer were manipulated. These factors can be divided into factors inherent to the subjects, inherent to the languages, or inherent to the grammatical structure or task. Language proficiency (and AoA) is a factor that is inherent to the subjects. Similarity of the languages, transparency, complexity, and symmetry of the gender systems are factors inherent to the languages. Agreement distance, online vs. offline tasks, and production vs. comprehension tasks are factors inherent to the structure or task tested, as well as Cognate Status which was manipulated in Experiment 1.

Regarding these factors, the following insights on gender transfer were gained:

Language families and symmetry of gender systems: Experiment 1 showed that gender transfer is possible between two different language families with asymmetric gender systems. Transfer seemed to occur at least in the acquisition stage resulting in faulty representations, as shown in the offline task, and also in the error rates of the PNT. Experiment 2 confirmed that gender transfer can occur when gender systems are asymmetric and extended this finding to online processing. Future research has to answer the question whether transfer is any more difficult between different language families and/or between languages with asymmetric gender systems than within language families and/or between languages with symmetric gender systems.

Transparency and proficiency: In Experiment 1 it was found that transparency of the L2 strongly mediates gender transfer: no transfer occurred for L2 Spanish with a transparent gender system, but transfer occurred for L2 German with an intransparent gender system[2]. In addition, for L2 Spanish, error rates were affected by noun ending transparency. Experiment 2 showed that there is a trade-off between transparency and proficiency: L1 transfer occurred into an L2 with an extremely transparent system at very low stages of proficiency and was overcome at slightly higher stages of proficiency. As shown in the first experiment, this was not the case for L2 German with a complex gender system, where transfer effects persisted even at higher proficiency levels.

Transparency and task demands/agreement distance: Experiment 1 showed that German subjects experienced no transfer in the offline gender assignment task. However, despite the great transparency of the Spanish gender system, they exhibited tendencies towards transfer in the error rates of the PNT. This might be attributed to the difference in task demands. Experiment 2 showed that transfer effects are possible into an L2 with an even more transparent gender system if task demands and agreement distance are further increased. In the first experiment, on the other hand, Spanish subjects showed transfer effects even in an offline task at a low agreement distance, due to the intransparency of the German gender system. This shows that there is a trade-off between transparency and task demands or agreement distance. Furthermore, it is likely that there is also a trade-off between agreement distance and proficiency, since in Experiment 2 no transfer effects were observed for an apparently slightly higher proficient subgroup. Exploratory analyses of the offline gender assignment task also indicated a possible trade-off between transparency and frequency.

Hence, in language transfer in late bilinguals, there are (at least) three important factors that strongly interact: transparency, proficiency, and task demands or agreement distance. These factors are inherent to the gender system, to the subjects, and to the nature of the task or syntactic structure, respectively. The strength of gender transfer depends on a trade-off between these factors. A similar trade-off of these factors also occurs for general ease of L2 gender processing (cf. section 2.2).

Cognate Status: Contrary to well-established findings in the literature, no effects of Cognate Status were found in Experiment 1. This is possibly due to the fact that on average, cognates were longer in terms of letters and syllables, less transparent, and less frequent than noncognates, that way canceling out possible facilitation effects. In Experiment 2, only cognate words were used as

[2] However, note that at least in this experiment also the fact that German has a three-way gender system might have played a role. As stated before, the odds of making a correct guess are 33 % in German but 50 % in Spanish.

antecedents which might have increased the likelihood of observing language transfer. Note that the question of the role of Cognate Status in gender transfer might be related to the question of the role of language similarity. If languages are very similar, that is, at least from the same language family, and the L2 lexicon thus contains many cognates, transfer might arise more easily (cf. section 3.1 on transfer effects in the lexicon).

AoA effects: The present thesis did not directly test AoA effects but tested the mastery of and transfer effects within a grammatical structure that is generally considered difficult for late L2 learners (cf. chapter 2). Due to the difficulty the acquisition of grammatical gender constitutes for late bilinguals, this structure was thought to be especially appropriate for observing some of the consequences of late AoA. The present thesis showed that difficulties with grammatical gender and transfer effects occur in late bilinguals. However, these effects were mediated by transparency, proficiency, and task demands/agreement distance. My experiments demonstrated that subjects´ performance was almost native-like even at lower proficiency levels when the L2 structure or task was sufficiently simple. This was the case in the offline gender assignment task where also low-proficient German subjects performed almost at ceiling, especially for transparent nouns. In Experiment 2, a subgroup of participants exhibited a (probably) native-like P600. These results suggest that at least in certain tasks and for certain transparent gender systems, age effects might be overcome and that grammatical gender does not have to remain problematic for late bilinguals in all cases. This trade-off between AoA and proficiency is in line with studies discussed in section 1.4.1.

Considering the present results combined with earlier discussed findings from the literature, several factors that can increase or decrease the likelihood for observing gender transfer can be identified. These factors are summarized in Table 6.1.

Naturally, in addition to the factors experimentally tested and described here, there might be several other factors mediating L1 gender transfer. For instance, the importance of the similarity of languages and the symmetry of gender systems on gender transfer is not clear from the present results and neither from results reported in the literature so far. Furthermore, since it has been shown that success in L2 gender acquisition is influenced by whether the L1 has a gender system or not (cf. section 3.3.1), it is possible that the factor whether the L1 has a complex gender system or a simple gender system also plays a role in gender acquisition and transfer. Another factor certainly is language mode/language environment and L1 use. It might be more difficult to prevent (gender) interference from L1 if the native language is still used a lot, which has, for example, been shown to be the case for L1 influences in L2 pronunciation, that is, foreign accent (Piske et al., 2001).

		Factors	**Type of factor**
Probability of gender transfer	**High**	High AoA	Subject factors
		Low proficiency	
		L2 gender system complex/intransparent	L2 factor
		High task demands (online/production)	Task factors
		High agreement distance	
	Low	Low AoA	Subject

		High proficiency	factors
		L2 gender system simple/transparent	L2 factor
		Low task demands (offline/comprehension)	Task factors
		Low agreement distance	

Table 6.1 Summary of factors increasing and decreasing the probability of gender transfer. Factors are either inherent to the subjects, to the L2, or to the task.

Regarding the models discussed in section 1.5 – the Declarative-Procedural (DP) Model and the Competition Model –, my findings are partially in line with their predictions. Both models predict that L2 learners do not show native-like performance in L2 grammar at low proficiency levels, but that this disadvantage might be overcome at higher proficiency stages. This was shown in Experiment 2 and possibly speaks for a greater reliance on procedural structures at greater proficiency levels. However, in response to the morphosyntactic violation, only a P600 and no LAN was found in the higher-proficient subgroup, which probably implies that this group was not quite native-like yet[3]. According to previous results reported in the literature, finding a LAN at higher proficiencies would have provided clearer evidence for the reliance on procedural structures in high-proficient learners. Finding an N400, on the other hand, would have provided clear evidence for the reliance on declarative structures in low-proficient learners. The DP Model also predicted difficulties at greater agreement distances, which was true for the lower-proficient subgroup in Experiment 2 who showed transfer in the error rates and no sensitivity to L2 violations in the ERPs.

The predictions of the Competition Model regarding language transfer were largely borne out. The languages used in the present experiments differed in gender congruency across conditions. In cases of gender incongruency, evidence for gender transfer was provided in both experiments. Moreover, in Experiment 2, the predictions regarding the strength of transfer for the different conditions (incongruent condition < correct condition < pseudocongruent condition) because of cue summation (Hernandez et al., 2005) was fulfilled in the error rates but not in the ERP results. The first experiment further showed that transparency cues are important which is in line with the importance of cue validity postulated by the Competition Model. In addition, the second experiment provided a new finding, showing that transfer occurs even under circumstances where the L1 feature does not exist in L2. According to my knowledge, for such a case, no predictions regarding transfer are specified in the Competition Model. Hence, the Competition Model needs to be extended in order to comprise an explanation for transfer of structures unique to L1.

As already mentioned in section 1.5, none of the models can account for all the phenomena and difficulties observed in L2 processing. The DP Model so far fails to provide an explanation for transfer effects and effects of transparency. The Competition Model postulates greater entrenchment for certain language arenas but lacks a more specific explanation for the great difficulty of grammatical gender in L2 learning. The reason for difficulties at greater agreement distances in L2 learners is also not accounted for by the Competition Model.

[3] Since no native control group was included, it is not known whether native speakers would have exhibited a LAN in response to the investigated morphosyntactic violation.

The claim made by the DP Model, that regular verbs are computed and irregular verbs are stored in declarative memory in native speakers, would be interesting to verify for the processing of nouns that are regular or irregular regarding their gender endings. Cubelli et al. (2005) postulate in their proposal of the “double selection mechanism” (p. 53; cf. also section 3.3.2) that regular or transparent nouns are computed. According to the DP Model, they should therefore be computed by the procedural system in native speakers, different from irregular nouns. Another question would be whether in late bilinguals these regular nouns are also processed in the declarative memory system, as claimed for regular verbs.

The description of the factors shown to mediate L1 gender transfer has an influence on the structure of the two gender systems of a bilingual (cf. the “gender-integrated” (p. 182) and the “language-autonomy” (p. 183) view (Costa et al., 2003) described in section 3.3.2). The fact that the occurrence of gender interference depends on the influence of other factors than factors inherent to the subjects or languages in question, such as task demands or agreement distance, supports the notion of an integrated bilingual gender system. As a consequence, gender transfer will potentially always arise in late bilinguals if task demands or agreement distance are only high enough. Hence, the gender systems of at least late bilinguals must interact. In cases where gender transfer does not arise, it is possible that task demands or agreement distance are just too low for transfer effects to become visible. As a matter of fact, it would not be feasible to postulate that the two gender systems are autonomous under certain circumstances (i.e., when task demands and agreement distance are low) but not under other circumstances (i.e., when task demands and agreement distance are high). Interference is just more successfully prevented under conditions of lower task demands, when more working memory resources are available. The prevention of interference (or the suppression of the influence of the non-target language) seems to be an ability early bilinguals are especially good at (e.g., Bialystok, 2005, 2006) and which might be affected by the amount of practice and working memory resources. It is possible that gender transfer only occurs at lower proficiency levels. However, in the literature, difficulties with L2 gender and transfer effects have been found also in high-proficient bilinguals. Therefore, it seems likely that under highly demanding circumstances (e.g., production tasks under time pressure requiring the calculation of long-distance agreements) interference could still be observed at higher proficiency levels. It is an interesting question whether under extremely high task demands L1 gender interference could be observed in very high-proficient late bilinguals or perhaps even early bilinguals.

Furthermore, regarding the acquisition-based vs. online account for gender transfer effects (Lemhöfer, Spalek, & Schriefers, 2008), my results point towards a shared gender system with strong L1 influences in the acquisition phase, leading to L2 gender representations which are parasitic on L1 (Experiment 1), as well as online transfer effects (Experiment 2). Because of the lack of a grammatical gender system in English, the online gender transfer effects found in the second experiment cannot be attributed to incorrect representations due to faulty acquisition. Hence, the results of the second experiment suggest that the two gender systems of a bilingual also interact in online processing.

In the future, it might be interesting to investigate differences in gender transfer between the incongruent condition and the incongruent neutral condition (cf. Experiment 1, chapter 4). The incongruent neutral condition consisted of nouns that are neutral in German and masculine or feminine in Spanish. The incongruent condition consisted of nouns that are feminine in German and masculine in Spanish and vice versa. Exploratory analysis of the data obtained in Experiment 1 raises the possibility that interference effects are stronger for the incongruent neutral condition than for the incongruent condition. The difference in error rates between the incongruent and incongruent neutral condition never reached significance. But when looking at the results it seemed that across tasks, transfer was stronger for incongruent neutral condition. In the error rates of the PNT (cf. section 4.3.3), error rates for the bilingual German group were 2.1 % in the incongruent condition and 2.6% incongruent neutral condition. For the bilingual Spanish group, error rates were 10.5 % in the incongruent condition and 12.6 % in the incongruent neutral condition. In the offline gender assignment task (cf. section 4.5), error rates for the bilingual German group were 2.7 % in the incongruent condition and 3.2 % in the incongruent neutral condition. In the bilingual Spanish group, error rates were 25.7 % incongruent condition and 28.9 % incongruent neutral condition. Hence, because of the lack of neutral gender in Spanish, the incongruent neutral condition might be even more prone to transfer effects than the incongruent condition .

Another research suggestion concerns the influence of the strength of “perceived gender” on gender transfer. Studies have shown that grammatical gender also influences gender at a conceptual level, even for inanimate objects that do not have biological gender (Konishi, 1993; Kurinski & Sera, 2011; Sera et al., 2002). Objects with masculine grammatical gender are perceived as more masculine than objects with feminine grammatical objects and are attributed more masculine characteristics, and vice versa. Therefore, I hypothesized that gender transfer might also be influenced by this “perceived gender” to a certain extent. In Experiment 2, perceived gender ratings were collected for the nouns used in the stimulus sentences. Subjects indicated the strength of perceived gender on a scale from 0 to 10 (0 = very masculine, 10 = very feminine; cf. section 5.2.2). Correlational analyses showed that there was a significant correlation between strength of perceived gender and error rate in the correct condition ($\rho = 0.281$, N =1.237, $p < 0.001$) and pseudocongruent condition ($\rho = 0.240$, N = 1237, $p < 0.001$), but not in the incongruent condition ($\rho = -0.002$, N = 1237, $p = 0.952$). Note that gender transfer was observed in the correct and pseudocongruent condition but not in the incongruent condition. Thus, it seems likely that L1 gender transfer is also influenced by the strength of perceived gender. The importance of the strength of perceived gender could be interesting to further investigate in future studies as it might have consequences for the acquisition of grammatical gender. It might be easier to acquire the correct grammatical gender of an L2 noun if perceived gender coincides with the L2 grammatical gender and if the perceived gender is very strong.

General conclusions

The aim of the present thesis was to describe the circumstances under which gender transfer occurs in L2 learning. The present thesis showed that L1 gender transfer is mediated by L2 characteristics, such as transparency. The influence of L2 characteristics has so far received little attention in research on transfer. Moreover, it was shown that transfer of L1 features that are not present in L2 is nevertheless possible. Another important finding is that in gender transfer, there is a trade-off between various factors such as proficiency, transparency of the L2 gender system, and agreement distance and task demands.

Nonetheless, as already remarked in the discussion of the offline gender assignment task, not all errors in L2 learning and not all difficulties late L2 learners experience can be explained by L1 (gender) transfer. Otherwise performance would be always at ceiling in cases of congruency. Success in late L2 acquisition is constrained by various factors and other possible reasons for the difficulties late learners experience in L2 acquisition have to be investigated as well. Yet, the present results once again show that the two gender systems or language systems of a (late) bilingual are not completely separate but influence each other, at various stages of proficiency. L1 (gender) transfer is a problem in late L2 acquisition that has to be taken seriously. Especially at a more advanced level, faulty morphosyntactic agreement will certainly impair comprehension. Ideally, teaching and learning methods that serve to minimize L1 interference (and at the same time possibly enhancing positive L1 transfer) should be developed.

7. References

Alario, F.-X., Ayora, P., Costa, A., & Melinger, A. (2008). Grammatical and nongrammatical contributions to closed-class word selection. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(4), 960–81.

Alario, F.-X., & Caramazza, A. (2002). The production of determiners: Evidence from French. *Cognition*, *82*(3), 179–223.

Alario, F.-X., Ferrand, L., Laganaro, M., New, B., Frauenfelder, U. H., & Segui, J. (2004). Predictors of picture naming speed. *Behavior Research Methods*, *36*(1), 140–55.

Alario, F.-X., Matos, R. E., & Segui, J. (2004). Gender congruency effects in picture naming. *Acta Psychologica*, *117*(2), 185–204.

Alderson, J. C. (2006). *Diagnosing foreign language proficiency: The interface between learning and assessment* (p. 284). London: Continuum.

Alderson, J. C., & Huhta, A. (2005). The development of a suite of computer-based diagnostic tests based on the Common European Framework. *Language Testing*, *22*(3), 301–320.

Altarriba, J., & Gianico, J. L. (2003). Lexical ambiguity resolution across languages: A theoretical and empirical review. *Experimental Psychology*, *50*(3), 159–170.

Antón-Méndez, I. (2010). Gender bender: Gender errors in L2 pronoun production. *Journal of Psycholinguistic Research*, *39*(2), 119–39.

Ardal, S., Donald, M. W., Meuter, R., Muldrew, S., & Luce, M. (1990). Brain responses to semantic incongruity in bilinguals. *Brain and Language*, *39*(2), 187–205.

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). The CELEX lexical database (Release 2) [CD-ROM]. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania (distributor).

Barber, H. A., & Carreiras, M. (2005). Grammatical gender and number agreement in Spanish: An ERP comparison. *Journal of Cognitive Neuroscience*, *17*(1), 137–53.

Barcelona European Council. (2002) (Vol. 35, pp. 0–72).

Barto-Sisamout, K., Nicol, J., Witzel, J., & Witzel, N. (2009). Transfer effects in bilingual sentence processing. *Arizona Working Papers in SLA & Teaching*, *16*, 1–26.

Bates, E. A., D’Amico, S., Jacobsen, T., Székely, A., Andonova, E., Devescovi, A., … Tzeng, O. (2003). Timed picture naming in seven languages. *Psychonomic Bulletin & Review*, *10*(2), 344–80.

Bates, E. A., Devescovi, A., Hernandez, A. E., & Pizzamiglio, L. (1996). Gender priming in Italian. *Perception & Psychophysics*, *58*(7), 992–1004.

Bates, E. A., & MacWhinney, B. (1987). Competition, variation, and language learning. In B. MacWhinney (Ed.), *Mechanisms of language acquisiton* (pp. 157–194). Hillsdale, NJ: Lawrence Erlbaum Associates.

Bates, E., Devescovi, A., Pizzamiglio, L., D’Amico, S., & Hernandez, A. (1995). Gender and lexical access in Italian. *Perception & Psychophysics*, *57*(6), 847–62.

Bergen, J. J. (1978). A simplified approach for teaching the gender of Spanish nouns. *Hispania*, *61*(4), 865–876.

Bhatia, T. K., & Ritchie, W. C. (2006). Introduction. In T. K. Bhatia & W. C. Ritchie (Eds.), *The Handbook of Bilingualism* (pp. 1–2). Oxford: Blackwell Publishing.

Bialystok, E. (2005). Consequences of bilingualism for cognitive development. In J. F. Kroll & A. De Groot (Eds.), *Handbook of Bilingualism: Psycholinguistic Approaches* (pp. 417–432).

Bialystok, E. (2006). The impact of bilingualism on language and literacy development. In T. K. Bhatia & W. C. Ritchie (Eds.), *The Handbook of Bilingualism* (pp. 577–601). Oxford: Blackwell Publishing.

Biemann, C., Bordag, S., Heyer, G., & Quasthoff, U. (2004). Language-independent methods for compiling monolingual lexical data. In *Proceedings of CIcLING, LNCS 2945.* (pp. 1–12).

Birdsong, D. (1992). Ultimate attainment in second language acquisition. *Language*, *68*(4), 706–755.

Birdsong, D. (2005a). Nativelikeness and non-nativelikeness in L2A research. *IRAL - International Review of Applied Linguistics in Language Teaching*, *43*, 319–328.

Birdsong, D. (2005b). Interpreting age effects in second language acquisition. In J. F. Kroll & A. De Groot (Eds.), *Handbook of Bilingualism: Psycholinguistic Approaches* (pp. 109–127). Oxford: Oxford University Press.

Birdsong, D. (2006). Age and second language acquisition and processing: A selective overview. *Language Learning*, *56*, 9–49.

Birdsong, D., & Molis, M. (2001). On the evidence for maturational constraints in second-language acquisition. *Journal of Memory and Language*, *44*(2), 235–249.

Blom, E., Polisenska, D., & Weerman, F. (2008). Articles, adjectives and age of onset: The acquisition of Dutch grammatical gender. *Second Language Research*, *24*(3), 297–331.

Bloomfield, L. (1933). *Language* (p. 564). New York: Holt.

Böhringer, A. (2000). *Langenscheidts Kurzgrammatik Spanisch* (p. 80). Berlin: Langenscheidt.

Bongaerts, T., van Summeren, C., Planken, B., & Schils, E. (1997). Age and ultimate attainment in the pronunciation of a foreign language. *Studies in Second Language Acquisition*, *19*, 447–465.

Bordag, D., Opitz, A., & Pechmann, T. (2006). Gender processing in first and second languages: The role of noun termination. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(5), 1090–101.

Bordag, D., & Pechmann, T. (2007). Factors influencing L2 gender processing. *Bilingualism: Language and Cognition*, *10*(03), 299–314.

Bordag, D., & Pechmann, T. (2008). Grammatical gender in translation. *Second Language Research*, *24*(2), 139–166.

Bowden, H. W., Gelfand, M. P., Sanz, C., & Ullman, M. T. (2010). Verbal inflectional morphology in L1 and L2 Spanish: A frequency effects study examining storage versus composition. *Language Learning*, *60*(1), 44–87.

Brinkmann, H. (1962). *Die deutsche Sprache. Gestalt und Leistung.* (p. 654). Düsseldorf: Schwann.

Caramazza, A. (1997). How many levels of processing are rhere in lexical access? *Cognitive Neuropsychology*, *14*(1), 177–208.

Caramazza, A., & Miozzo, M. (1997). The relation between syntactic and phonological knowledge in lexical access: Evidence from the "tip-of-the-tongue" phenomenon. *Cognition*, *64*(3), 309–43.

Caselli, M. C., Leonard, L. B., Volterra, V., & Campagnoli, M. G. (1993). Toward mastery of Italian morphology: A cross-sectional study. *Journal of Child Language*, *20*(02), 377–393.

Chen, L., Shu, H., Liu, Y., Zhao, J., & Li, P. (2007). ERP signatures of subject–verb agreement in L2 learning. *Bilingualism: Language and Cognition*, *10*(02), 161.

Clahsen, H., & Almazan, M. (1998). Syntax and morphology in Williams syndrome. *Cognition*, *68*(3), 167–98.

Clahsen, H., Eisenbeiss, S., & Sonnenstuhl-Henning, I. (1997). Morphological structure and the processing of inflected words. *Theoretical Linguistics*, *23*(3).

Clahsen, H., & Felser, C. (2006a). Authors´response. Continuity and shallow structures in language processing. *Applied Psycholinguistics*, *27*, 107–126.

Clahsen, H., & Felser, C. (2006b). How native-like is non-native language processing? *Trends in Cognitive Sciences*, *10*(12), 564–70.

Clahsen, H., & Felser, C. (2006c). Grammatical processing in language learners. *Applied Psycholinguistics*, *27*, 3–42.

Comrie, B. (1999). Grammatical gender systems: A linguist's assessment. *Journal of Psycholinguistic Research*, *28*(5), 457–466.

Conklin, K., Dijkstra, T., & van Heuven, W. (2007). Influence of Dutch gender on discourse processing in English: Evidence from eyetracking. In *ESCoP2007*. Marseille, France.

Coppieters, R. (1987). Competence differences between native and near-native speakers. *Language*, *63*(3), 544–573.

Corbett, G. G. (1991). *Gender* (Vol. 1). Cambridge: Cambridge University Press.

Costa, A., Caramazza, A., & Sebastián-Gallés, N. (2000). The cognate facilitation effect: Implications for models of lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(5), 1283–96.

Costa, A., Kovacic, D., Franck, J., & Caramazza, A. (2003). On the autonomy of the grammatical gender systems of the two languages of a bilingual. *Bilingualism: Language and Cognition*, *6*(3), 181–200.

Costa, A., & Santesteban, M. (2004). Bilingual word perception and production: Two sides of the same coin? *Trends in Cognitive Sciences*, *8*(6), 253.

Costa, A., Sebastián-Gallés, N., Miozzo, M., & Caramazza, A. (1999). The gender congruity effect: Evidence from Spanish and Catalan. *Language and Cognitive Processes*, *14*(4), 381–391.

Coulson, S., & Kutas, M. (2001). Getting it: Human event-related brain response to jokes in good and poor comprehenders. *Neuroscience letters*, *316*(2), 71–4.

Council of Europe. (2001). *The Common European Framework of Reference for Languages*. *Europe* (p. 264). Cambridge: Cambridge University Press.

Cubelli, R., Lotto, L., Paolieri, D., Girelli, M., & Job, R. (2005). Grammatical gender is selected in bare noun production: Evidence from the picture–word interference paradigm. *Journal of Memory and Language*, *53*(1), 42–59.

Cubelli, R., & Paolieri, D. (2008). The processing of grammatical gender as abstract lexical feature. In L. Arcuri, P. Boscolo, & F. Peressotti (Eds.), *Language and cognition: a long story. Festschrift in honour of Ino Flores D' Arcais* (pp. 73–86). Padova, Italy: Cleup.

Cuetos, F., & Mitchell, D. C. (1988). Cross-linguistic differences in parsing: Restrictions on the use of the Late Closure strategy in Spanish. *Cognition*, *30*, 73–105.

Curran, T., & Dien, J. (2003). Differentiating amodal familiarity from modality-specific memory processes: An ERP study. *Psychophysiology*, *40*(6), 979–988.

Dahan, D., Swingley, D., Tanenhaus, M. K., & Magnuson, J. S. (2000). Linguistic gender and spoken-word recognition in French. *Journal of Memory and Language*, *42*(4), 465–480.

Dambacher, M., Kliegl, R., Hofmann, M., & Jacobs, A. M. (2006). Frequency and predictability effects on event-related potentials during reading. *Brain Research*, *1084*(1), 89–103.

Davidson, D. J., & Indefrey, P. (2009). An event-related potential study on changes of violation and error responses during morphosyntactic learning. *Journal of Cognitive Neuroscience*, *21*(3), 433–46.

Davis, C. J., & Perea, M. (2005). BuscaPalabras: A program for deriving orthographic and phonological neighborhood statistics and other psycholinguistic indices in Spanish. *Behavior Research Methods*, *37*(4), 665–71.

De Bruyne, J. (2002). *Spanische Grammatik* (p. 663). Tübingen: Max Niemeyer Verlag.

De Diego Balaguer, R., Sebastián-Gallés, N., Díaz, B., & Rodríguez-Fornells, A. (2005). Morphological processing in early bilinguals: An ERP study of regular and irregular verb processing. *Cognitive Brain Research*, *25*(1), 312–27.

Degani, T., Prior, A., & Tokowicz, N. (2011). Bidirectional transfer: The effect of sharing a translation. *Journal of Cognitive Psychology*, *23*(1), 18–28.

DeKeyser, R. M. (2005). What makes learning second-language grammar difficult? A review of issues. *Language Learning*, *55*(S1), 1–25.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, *93*(3), 283–321.

Desmet, T., & Declercq, M. (2006). Cross-linguistic priming of syntactic hierarchical configuration information. *Journal of Memory and Language*, *54*(4), 610–632.

Dewaele, J.-M., & Furnham, A. (1999). Extraversion: The unloved variable in applied linguistic research. *Language Learning*, *49*(3), 509–544.

Dewaele, J.-M., & Furnham, A. (2000). Personality and speech production: A pilot study of second language learners. *Personality and Individual Differences*, *28*, 355–365.

Dewaele, J.-M., & Véronique, D. (2001). Gender assignment and gender agreement in advanced French interlanguage: A cross-sectional study. *Bilingualism: Language and Cognition*, *4*(03), 275–297.

Dijkstra, T. (2005). Bilingual visual word recognition and lexical access. In J. F. Kroll & A. M. B. De Groot (Eds.), *Handbook of Bilingualism: Psycholinguistic Approaches* (pp. 179–201). Oxford: Oxford University Press.

Dijkstra, T., Grainger, J., & van Heuven, W. J. B. (1999). Recognition of cognates and interlingual homographs: the neglected role of phonology. *Journal of Memory and Language*, *41*(4), 496–518. 54

Dijkstra, T., & van Heuven, W. J. B. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition*, *5*(03).

Dixon, R. M. W. (1968). Noun classes. *Lingua*, *21*, 104–125.

Duenas de Haensch, M. (1999). *Langenscheidt Grundwortschatz Spanisch* (p. 322). Berlin: Langenscheidt.

Dussias, P. E. (2003). Syntactic ambiguity resolution in L2 learners: Some effects of bilinguality on L1 and L2 processing strategies. *Studies in Second Language Acquisition*, *25*, 529–557.

Dussias, P. E., & Sagarra, N. (2007). The effect of exposure on syntactic parsing in Spanish–English bilinguals. *Bilingualism: Language and Cognition*, *10*(01), 101–116.

Eisenberg, P. (2013). *Grundriss der deutschen Grammatik 2: Der Satz* (pp. 1–535). Stuttgart: Verlag J.B. Metzler.

Eisenstein, M. (2008). Native reactions to non-native speech: A review of empirical research. *Studies in Second Language Acquisition*, *5*(02), 160.

Elston-Güttler, K. E., Paulmann, S., & Kotz, S. A. (2005). Who's in control? Proficiency and L1 influence on L2 processing. *Journal of Cognitive Neuroscience*, *17*(10), 1593–610.

Elston-Güttler, K. E., & Williams, J. N. (2008). First language polysemy affects second language meaning interpretation: Evidence for activation of first language concepts during second language reading. *Second Language Research*, *24*(2), 167–187.

Embick, D., & Marantz, A. (2005). Cognitive neuroscience and the English past tense: Comments on the paper by Ullman et al. *Brain and Language*, *93*, 243–247.

Evans, K. M., & Federmeier, K. D. (2007). The memory that's right and the memory that's left: Event-related potentials reveal hemispheric asymmetries in the encoding and retention of verbal information. *Neuropsychologia*, *45*(8), 1777–90.

Federmeier, K. D., & Kutas, M. (2002). Picture the difference: Electrophysiological investigations of picture processing in the two cerebral hemispheres. *Neuropsychologia*, *40*(7), 730–47.

Federmeier, K. D., Mai, H., & Kutas, M. (2005). Both sides get the point: Hemispheric sensitivities to sentential constraint. *Memory & Cognition*, *33*(5), 871–886.

Felser, C., Roberts, L., Marinis, T., & Gross, R. (2003). The processing of ambiguous sentences by first and second language learners of English. *Applied Psycholinguistics*, *24*(03), 453–489.

Fernald, A., Perfors, A., & Marchman, V. A. (2006). Picking up speed in understanding: Speech processing efficiency and vocabulary growth across the 2nd year. *Developmental Psychology*, *42*, 98–116.

Flege, J. E., Bohn, O.-S., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics*, *25*, 437–470.

Flege, J. E., Yeni-Komshian, G. H., & Liu, S. (1999). Age constraints on second-language acquisition. *Journal of Memory and Language*, *41*(1), 78–104.

Foote, R. (2011). Integrated knowledge of agreement in early and late English–Spanish bilinguals. *Applied Psycholinguistics*, *32*(01), 187–220.

Foucart, A., & Frenck-Mestre, C. (2011). Grammatical gender processing in L2: Electrophysiological evidence of the effect of L1–L2 syntactic similarity. *Bilingualism: Language and Cognition*, *14*(03), 379–399.

Foucart, A., & Frenck-Mestre, C. (2012). Can late L2 learners acquire new grammatical features? Evidence from ERPs and eye-tracking. *Journal of Memory and Language*, *66*(1), 226–248.

Franceschina, F. (2001). Morphological or syntactic deficits in near-native speakers? An assessment of some current proposals. *Second Language Research*, *17*(3), 213–247.

Franceschina, F. (2005). *Fossilized second language grammars. The acquisition of grammatical gender.* (Vol. 4, p. 288). Amsterdam: John Benjamins.

Frenck-Mestre, C. (2002). An on-line look at sentence processing in the second language. In R. R. Heredia & J. Altarriba (Eds.), *Bilingual Sentence Processing* (pp. 217–236).

Frenck-Mestre, C., Foucart, A., Carrasco, H., & Herschensohn, J. (2009). Processing of grammatical gender in French as a first and second language. Evidence from ERPs. *EUROSLA Yearbook*, *9*(1), 76–106.

Frenck-Mestre, C., Osterhout, L., McLaughlin, J., & Foucart, A. (2008). The effect of phonological realization of inflectional morphology on verbal agreement in French: Evidence from ERPs. *Acta Psychologica*, *128*(3), 528–36.

Frenck-Mestre, C., & Pynte, J. (1997). Syntactic ambiguity resolution while reading in second and native languages. *The Quarterly Journal of Experimental Psychology*, *50A*(1), 119–148.

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences*, *6*(2), 78–84.

Friederici, A. D., Steinhauer, K., & Pfeifer, E. (2002). Brain signatures of artificial language processing: Evidence challenging the critical period hypothesis. *Proceedings of the National Academy of Sciences*, *99*(1), 529–534.

Fromkin, V., Krashen, S. D., Curtiss, S., Rigler, D., & Rigler, M. (1974). The development of language in genie: A case of language acquisition beyond the “critical period.” *Brain and Language*, *1*, 81–107.

Ganushchak, L. Y., Verdonschot, R. G., & Schiller, N. O. (2011). When leaf becomes neuter: Event-related potential evidence for grammatical gender transfer in bilingualism. *NeuroReport*, *22*(3), 106–10.

Gillon Dowens, M., Guo, T., Guo, J., Barber, H., & Carreiras, M. (2011). Gender and number processing in Chinese learners of Spanish – Evidence from Event Related Potentials. *Neuropsychologia*, *49*(7), 1651–1659.

Gillon Dowens, M., Vergara, M., Barber, H. A., & Carreiras, M. (2010). Morphosyntactic processing in late second-language learners. *Journal of Cognitive Neuroscience*, *22*(8), 1870–1887.

Götze, L., & Hess-Lüttich, E. W. B. (1999). *Grammatik der deutschen Sprache: Sprachsystem und Sprachgebrauch* (p. 702). Gütersloh: Bertelsmann Lexikon Verlag.

Gratton, G., Coles, M. G. H., & Donchin, E. (1983). A new method for off-line removal of ocular artifact. *Electroencephalography and Clinical Neurophysiology*, *55*, 468–484.

Greenhouse-Geisser. (1959). On methods in the analysis of profile data. *Psychometrika*, *24*, 95–111.

Grimshaw, G. M., Adelstein, A., Bryden, M. P., & MacKinnon, G. E. (1998). First-language acquisition in adolescence: Evidence for a critical period for verbal language development. *Brain and Language*, *63*(2), 237–55.

Groot, A. M. B. De, & van Hell, J. G. (2005). The learning of foreign language vocabulary. In J. F. Kroll & A. M. B. De Groot (Eds.), *Handbook of Bilingualism: Psycholinguistic Approaches* (pp. 9–29). Oxford: Oxford University Press.

Grosjean, F. (1989). Neurolinguists, beware! The bilingual is not two monolinguals in one person. *Brain and Language*, *36*, 3–15.

Grosjean, F. (1998a). Transfer and language mode. *Bilingualism: Language and Cognition*, *1*(3), 175–176.

Grosjean, F. (1998b). Studying bilinguals: Methodological and conceptual issues. *Bilingualism: Language and Cognition*, *1*(2), 131–149.

Grosjean, F. (1999). The bilingual's language modes. In J. L. Nicol (Ed.), *One Mind, Two Languages: Bilingual Language Processing.* (pp. 1–22). Oxford: Blackwell Publishing.

Grüter, T., Lew-Williams, C., & Fernald, A. (2012). Grammatical gender in L2: A production or a real-time processing problem? *Second Language Research*, *28*(2), 191–215.

Guillelmon, D., & Grosjean, F. (2001). The gender marking effect in spoken word recognition: The case of bilinguals. *Memory & Cognition*, *29*(3), 503–11.

Guo, J., Guo, T., Yan, Y., Jiang, N., & Peng, D. (2009). ERP evidence for different strategies employed by native speakers and L2 learners in sentence processing. *Journal of Neurolinguistics*, *22*(2), 123–134.

Gupta, P., & MacWhinney, B. (1992). Integrating cateogry acquisition with inflectional marking: A model of the German nominal system. In *Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society* (pp. 253–258). Hillsdale, NJ: Lawrence Erlbaum Associates.

Gurjanov, M., Lukatela, G., Lukatela, K., Savic, M., & Turvey, M. (1985). Grammatical priming of inflected nouns by the gender of possessive adjectives. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*(4), 692–701.

Hahne, A. (2001). What's different in second-language processing? Evidence from event-related brain potentials. *Journal of Psycholinguistic Research*, *30*(3), 251–266.

Hahne, A., & Friederici, A. D. (2001). Processing a second language: Late learners' comprehension mechanisms as revealed by event-related brain potentials. *Bilingualism: Language and Cognition*, *4*(02), 123–141.

Hahne, A., Mueller, J. L., & Clahsen, H. (2006). Morphological processing in a second language: Behavioral and event-related brain potential evidence for storage and decomposition. *Journal of Cognitive Neuroscience*, *18*(1), 121–34.

Hakuta, K., Bialystok, E., & Wiley, E. (2003). Critical evidence: A test of the critical-period hypothesis for second-language acquisition. *Psychological Science*, *14*(1), 31–38.

Hammer, A., Jansma, B. M., Lamers, M., & Münte, T. F. (2005). Pronominal reference in sentences about persons or things: An electrophysiological approach. *Journal of Cognitive Neuroscience*, *17*(2), 227–39.

Hartsuiker, R. J., Pickering, M. J., & Veltkamp, E. (2004). Is syntax separate or shared between languages? *Psychological Science*, *15*(6), 409–414.

Hauk, O., & Pulvermüller, F. (2004). Effects of word length and frequency on the human event-related potential. *Clinical Neurophysiology*, *115*(5), 1090–103.

Hawkins, R., & Chan, Y. (1997). The partial availability of Universal Grammar in second language acquisition: The "failed functional features hypothesis." *Second Language Research*, *13*(3), 187–226.

Hernandez, A. E., Bates, E. A., & Avila, L. X. (1994). On-line sentence interpretation in Spanish–English bilinguals: What does it mean to be “in between”? *Applied Psycholinguistics*, *15*(04), 417.

Hernandez, A. E., Hofmann, J., & Kotz, S. A. (2007). Age of acquisition modulates neural activity for both regular and irregular syntactic functions. *Neuroimage*, *36*(3), 912–923.

Hernandez, A. E., Kotz, S. A., Hofmann, J., Valentin, V. V, Dapretto, M., & Susan, Y. (2004). The neural correlates of grammatical gender decisions in Spanish. *NeuroReport*, *15*(5), 63–66.

Hernandez, A. E., & Li, P. (2007). Age of acquisition: Its neural and computational mechanisms. *Psychological Bulletin*, *133*(4), 638–50.

Hernandez, A. E., Li, P., & MacWhinney, B. (2005). The emergence of competing modules in bilingualism. *Trends in Cognitive Sciences*, *9*(5), 220–5.

Hillyard, S. A., & Münte, T. F. (1984). Selective attention to color and location: An analysis with event-related brain potentials. *Perception & Psychophysics*, *36*(2), 185–98.

Hoberg, U., & Hoberg, R. (2009). *Der kleine Duden. Deutsche Grammatik* (p. 451). Mannheim: Dudenverlag.

Holcomb, P. J., Coffey, S. A., & Neville, H. J. (1992). Visual and auditory sentence processing: A developmental analysis using event-related brain potentials. *Developmental Neuropsychology*, *8*(2-3), 203–241.

Hosoda, M., Nguyen, L. T., & Stone-Romero, E. F. (2012). The effect of Hispanic accents on employment decisions. *Journal of Managerial Psychology*, *27*(4), 347–364.

Hosoda, M., & Stone-Romero, E. (2010). The effects of foreign accents on employment-related decisions. *Journal of Managerial Psychology*, *25*(2), 113–132.

Huang, Y. (2000). *Anaphora. A Cross-linguistic Study.* (p. 396). Oxford: Oxford University Press.

Hubel, D. H., & Wiesel, T. N. (1959). Receptive fields of single neurones in the cat´s striate cortex. *Journal of Physiology*, *148*, 574–591.

Hubel, D. H., & Wiesel, T. N. (1962). Receptive rields, binocular interaction and functional architecture in the cat´s visual cortex. *Journal of Physiology*, *160*, 106–154.

Hubert, M. D. (2011). Foreign language production and avoidance in US university Spanish-language education. *Language Learning*, *21*(2), 222–243.

Ioup, G., Boustagui, E., Tigi, M. El, & Moselle, M. (1994). Reexamining the critical period hypothesis. A case study of successful adult SLA in a naturalistic environment. *Studies in Second Language Acquisition*, *16*, 73–98.

Jackson, C. N., & Dussias, P. E. (2009). Transfer of sentence processing strategies: A comparison of L2 learners of Chinese and English. *Bilingualism: Language and Cognition*, *12*(1), 65–82.

Jacobsen, T. (1999). Effects of grammatical gender on picture and word naming: Evidence from German. *Journal of Psycholinguistic Research*, *28*(5), 499–514.

Jakubowicz, C., & Faussart, C. (1998). Gender agreement in the processing of spoken French. *Journal of Psycholinguistic Research*, *27*(6), 597–617.

Jiang, N. (2004). Morphological insensitivity in second language processing. *Applied Psycholinguistics*, *25*(04), 603–634.

Jiang, N. (2007). Selective integration of linguistic knowledge in adult second language learning. *Language Learning*, *57*(1), 1–33.

Jiang, X., Tan, Y., & Zhou, X. (2009). Processing the universal quantifier during sentence comprehension: ERP evidence. *Neuropsychologia*, *47*(8-9), 1799–815.

Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, *21*(1), 60–99.

Juffs, A. (1998). Main verb versus reduced relative clause ambiguity resolution in L2 sentence processing. *Language Learning*, *1*, 107–147.

Juffs, A. (2005). The influence of first language on the processing of wh-movement in English as a second language. *Second Language Research*, *21*(2), 121–151.

Kaan, E., & Swaab, T. Y. (2003). Repair, revision, and complexity in syntactic analysis: An electrophysiological differentiation. *Journal of cognitive neuroscience*, *15*(1), 98–110.

Kantola, L., & van Gompel, R. P. G. (2011). Between- and within-language priming is the same: Evidence for shared bilingual syntactic representations. *Memory & Cognition*, *39*(2), 276–90.

Keating, G. D. (2009). Sensitivity to violations of gender agreement in native and nonnative Spanish: An eye-movement investigation. *Language Learning*, *59*(3), 503–535.

Kempe, V., Brooks, P. J., & Kharkhurin, A. (2010). Cognitive predictors of generalization of Russian grammatical gender categories. *Language Learning*, *60*(1), 127–153.

Kess, J. F. (1992). *Psycholinguistics* (pp. 1–360). Amsterdam, Netherlands: John Benjamins Publishing Company.

Key Data on Teaching Languages at School in Europe. (2008).

Kilborn, K. (1989). Sentence processing in a second language: The timing of transfer. *Language and Speech*, *32*(1), 1–23.

Kleinmann, H. H. (1977). Avoidance behavior in adult second language acquisition. *Language Learning*, *27*(1), 93–107.

Konishi, T. (1993). The semantics of grammatical gender: A cross-cultural study. *Journal of Psycholinguistic Research*, *22*(5), 519–34.

Konrad Adenauer Stiftung. (2012). *Nigeria – Landesinformationen*. Retrieved November 07, 2012, from http://www.kas.de/nigeria/de/pages/3052/

Köpcke, K.-M. (1982). *Untersuchungen zum Genussystem der deutschen Gegenwartssprache* (p. 268). Tübingen: Niemeyer.

Köpcke, K.-M., & Zubin, D. A. (1983). Die kognitive Organisation der Genuszuweisung zu den einsilbigen Nomen der deutschen Gegenwartssprache/The cognitive organization of gender assignement of monosyllabic nouns in contemporary German. *Zeitschrift für germanistische Linguistik*, *11*, 166–182.

Kotz, S. A. (2009). A critical review of ERP and fMRI evidence on L2 syntactic processing. *Brain and Language*, *109*(2-3), 68–74.

Kroll, J. F., & Bogulski, C. A. (2013). Organization of the second language lexicon. *The Encyclopedia of Applied Linguistics*.

Kroll, J. F., & Dussias, P. E. (2006). The comprehension of words and sentences in two languages. In *The Handbook of Bilingualism* (pp. 169–200).

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language*, *33*, 149–174.

Kroll, J. F., & Tokowicz, N. (2005). Models of Bilingual Reresentation and Processing: Looking Back and to the Future. In *Handbook of Bilingualism: Psycholinguistic Approaches* (pp. 531–554). New York: Oxford University Press.

Kurinski, E., & Sera, M. D. (2011). Does learning Spanish grammatical gender change English-speaking adults' categorization of inanimate objects? *Bilingualism: Language and Cognition*, *14*(02), 203–220.

La Heij, W., Mak, P., Sander, J., & Willeboordse, E. (1998). The gender-congruency effect in picture-word tasks. *Psychological Research*, *61*(3), 209–219.

Lamers, M. J., Jansma, B. M., Hammer, A., & Münte, T. F. (2008). Differences in the processing of anaphoric reference between closely related languages: Neurophysiological evidence. *BMC neuroscience*, *9*, 55.

Lebrun, Y. (1980). Victor of Aveyron: A reappraisal in light of more recent cases of feral speech. *Language Sciences*, *2*(1), 32–43.

Lee, E.-J. (2001). Interlanguage development by two Korean speakers of English with a focus on temporality. *Language Learning*, *51*(4), 591–633.

Lemhöfer, K., Schriefers, H. J., & Hanique, I. (2010). Native language effects in learning second-language grammatical gender: A training study. *Acta Psychologica*, *135*(2), 150–8.

Lemhöfer, K., Spalek, K., & Schriefers, H. J. (2008). Cross-language effects of grammatical gender in bilingual word recognition and production. *Journal of Memory and Language*, *59*(3), 312–330.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation* (p. 566). Cambridge, MA: MIT Press.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*(1), 1–38; discussion 38–75.

Lew-Williams, C., & Fernald, A. (2007). Young children learning Spanish make rapid use of grammatical gender in spoken word recognition. *Psychological Science*, *18*(3), 193–8.

Lew-Williams, C., & Fernald, A. (2010). Real-time processing of gender-marked articles by native and non-native Spanish speakers. *Journal of Memory and Language*, *63*(4), 447–464.

Li, P., & MacWhinney, B. (2012). Competition Model. In C. A. Chapelle (Ed.), *The Encyclopedia of Applied Linguistics* (pp. 1–5). Blackwell Publishing.

Liu, H., Bates, E. A., & Li, P. (1992). Sentence interpretation in bilingual speakers of English and Chinese. *Applied Psycholinguistics*, *13*(04), 451–484.

Loebell, H., & Bock, K. (2003). Structural priming across languages. *Linguistics*, *41*(5), 791–824.

Lorenz, K. (1935). Der Kumpan in der Umwelt des Vogels. Der Artgenosse als auslösendes Moment sozialer Verhaltensweisen. *Journal für Ornithologie*, *2*, 137–213.

Luck, S. J. (2005). *An introduction to the event-related potential technique*. Cambridge, MA: The MIT Press.

Luck, S. J., & Hillyard, S. A. (1994). Electrophysiological correlates of feature analysis during visual search. *Psychophysiology*, *31*, 291 – 308.

MacWhinney, B. (1987a). *The Competition Model. Mechanisms of language acquisition.* Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.

MacWhinney, B. (1987b). Applying the Competition Model to bilingualism. *Applied Psycholinguistics*, *8*, 315–327.

MacWhinney, B. (1997). Second language acquisition and the Competition Model. In Annette M.B. de Groot & J. F. Kroll (Eds.), *Tutorials in Bilingualism: Psycholinguistic Perspectives*. Mahwah, NJ: Lawrence Erlbaum Associates.

MacWhinney, B. (2005a). A unified model of language acquisition. In Judith F; Kroll & A. DeGroot (Eds.), *Handbook of Bilingualism: Psycholinguistic Approaches* (pp. 49–67). Oxford: Oxford University Press.

MacWhinney, B. (2005b). Extending the Competition Model. *International Journal of Bilingualism*, *9*(1), 69–84.

MacWhinney, B. (2005c). Commentary on Ullman et al. *Brain and Language*, *93*(2), 239–242.

MacWhinney, B. (2008). A unified model. In *Handbook of Cognitive Linguistics and Second Language Acquisition.* (pp. 341–371). Mahwah, NJ: Lawrence Erlbaum Associates.

MacWhinney, B. (2011). The logic of the unified model. In S. Gass & A. Mackey (Eds.), *Handbook of Second Language Acquisition* (pp. 211–227). New York: Routledge.

MacWhinney, B., & Leinbach, J. (1991). Implementations are not conceptualizations: Revising the verb learning model. *Cognition*, *40*(1-2), 121–57.

Malaia, E., Wilbur, R. B., & Weber-Fox, C. M. (2009). ERP evidence for telicity effects on syntactic processing in garden-path sentences. *Brain and Language*, *108*(3), 145–58.

Marian, V., & Spivey, M. (2003a). Competing activation in bilingual language processing: Within- and between-language competition. *Bilingualism: Language and Cognition*, *6*(2), 97–115.

Marian, V., & Spivey, M. (2003b). Bilingual and monolingual processing of competing lexical items. *Applied Psycholinguistics*, *24*(02), 173–193.

Marinis, T., Roberts, L., Felser, C., & Clahsen, H. (2005). Gaps in second language sentence processing. *Studies in Second Language Acquisition*, *27*(01), 53–78.

Marler, P. (1991). Song-learning behavior: The interface with neuroethology. *Trends in Neurosciences*, *14*(5), 199–206.

Marler, P., & Peters, S. (1987). A sensitive period for song acquisition in the song sparrow, melospiza melodia: A case of age-limited learning. *Ethology*, *76*(2), 89–100.

Mayberry, R. I. (1993). First-Language acquisition after childhood differs from second-language acquisition: The case of American Sign Language. *Journal of Speech and Hearing Research*, *36*, 1258–1270.

Mayberry, R. I., & Eichen, E. B. (1991). The long-lasting advantage of learning sign language in childhood: Another look at the critical period for language acquisition. *Journal of Memory and Language*, *30*(4), 486–512.

Mayberry, R. I., & Lock, E. (2003). Age constraints on first versus second language acquisition: Evidence for linguistic plasticity and epigenesis. *Brain and Language*, *87*(3), 369–384.

McClelland, J. L., & Rumelhart, D. E. (1985). Distributed memory and the representation of general and specific information. *Journal of Experimental Psychology: General*, *114*(2), 159–97.

McDonald, J. L. (1987). Sentence interpretation in bilingual speakers of English and Dutch. *Applied Psycholinguistics*, *8*, 379–413.

McDonald, J. L. (2000). Grammaticality judgments in a second language: Influences of age of acquisition and native language. *Applied Psycholinguistics*, *21*(3), 395–423.

McDonald, J. L. (2006). Beyond the critical period: Processing-based explanations for poor grammaticality judgment performance by late second language learners. *Journal of Memory and Language*, *55*(3), 381–401.

McLaughlin, J., Osterhout, L., & Kim, A. (2004). Neural correlates of second-language word learning: Minimal instruction produces rapid change. *Nature Neuroscience*, *7*(7), 703–4.

McLaughlin, J., Tanner, D., Frenck-Mestre, C., Valentine, G., & Osterhout, L. (2010). Brain potentials reveal discrete stages of L2 grammatical learning. *Language Learning*, *60*(2), 123–150.

Midgley, K. J., Wicha, N. Y. Y., Holcomb, P. J., & Grainger, J. (2007). An electrophysiological study of gender agreement transfer in early language learners. In *Proceedings of the Annual Meeting of the Cognitive Neuroscience Society* (p. 170).

Miozzo, M., & Caramazza, A. (1999). The selection of determiners in noun phrase production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(4), 907–22.

Miozzo, M., Costa, A., & Caramazza, A. (2002). The absence of a gender congruency effect in romance languages: A matter of stimulus onset asynchrony? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(2), 388–391.

Montrul, S. (2001). First-language-constrained variability in the second-language acquisition of argument-structure-changing morphology with causative verbs. *Second Language Research*, *17*(2), 144–194.

Montrul, S., Foote, R., & Perpiñán, S. (2008). Gender agreement in adult second language learners and Spanish heritage speakers: The effects of age and context. *Language Learning*, *58*(3), 503–553.

Montrul, S., & Slabakova, R. (2003). Competence similarities between native and near-native speakers. *Studies in Second Language Acquisition*, *25*, 351–398.

Morales, L., Paolieri, D., & Bajo, T. (2011). Grammatical gender inhibition in bilinguals. *Frontiers in Psychology*, *2*, 1–9.

Morgan-Short, K., Finger, I., Grey, S., & Ullman, M. T. (2012). Second language processing shows increased native-like neural responses after months of no exposure. (E. A. Stamatakis, Ed.)*PLoS ONE*, *7*(3), 1–18.

Morgan-Short, K., Sanz, C., Steinhauer, K., & Ullman, M. T. (2010). Second language acquisition of gender agreement in explicit and implicit training conditions: An event-related potential study. *Language Learning*, *60*(1), 154–193.

Morgan-Short, K., Steinhauer, K., Sanz, C., & Ullman, M. T. (2012). Explicit and implicit second language training differentially affect the achievement of native-like brain activation patterns. *Journal of Cognitive Neuroscience*, *24*(4), 933–47.

Morgan-Short, K., & Ullman, M. T. (2011). The neurocognition of second language. In S. M. Gass & A. Mackey (Eds.), *The Handbook of Second Language Acquisition* (pp. 282–299). New York: Routledge.

Moriena, C., & Genschow, K. (2010). *Große Lerngrammatik Spanisch: Regeln, Anwendungsbeispiele, Tests* (p. 708). Ismaning: Hueber.

Moyer, A. (1999). Ultimate attainment in L2 phonology: The critical factors of age, motivation, and instruction. *Studies in Second Language Acquisition*, *21*(1), 81–108.

Mueller, J. L., Hahne, A., Fuji, Y., & Friederici, A. D. (2005). Native and nonnative speakers' processing of a miniature version of Japanese as revealed by ERPs. *Journal of Cognitive Neuroscience*, *17*(8), 1229–44.

Müller, O., & Hagoort, P. (2006). Access to lexical information in language comprehension: Semantics before syntax. *Journal of Cognitive Neuroscience*, *18*(1), 84–96.

Multilingualism: An Asset for Europe and a Shared Commitment (COM) 2008 566 final. (2008).

Myles, F. (1995). Interaction between linguistic theory and language processing in SLA. *Second Language Research*, *11*(3), 235–266.

Neville, H. J., Mills, D. L., & Lawson, D. S. (1992). Fractionating language: Different neural subsystems with different sensitive periods. *Cerebral Cortex*, *2*(3), 244–58.

Newman, A. J., Tremblay, A., Nichols, E. S., Neville, H. J., & Ullman, M. T. (2012). The influence of language proficiency on lexical semantic processing in native and late learners of English. *Journal of Cognitive Neuroscience*, *24*(5), 1205–1223.

Newman, A. J., Ullman, M. T., Pancheva, R., Waligura, D. L., & Neville, H. J. (2007). An ERP study of regular and irregular English past tense inflection. *NeuroImage*, *34*(1), 435–45.

Nigeria Embassy Berlin. (2012). *About Nigeria – General. Notes & Statistics.* Retrieved November 07, 2012, from http://www.nigeriaembassygermany.org/nigeria_general.htm

Nitschke, S., Kidd, E., & Serratrice, L. (2010). First language transfer and long-term structural priming in comprehension. *Language and Cognitive Processes*, *25*(1), 94–114.

Odlin, T. (1989). *Language transfer: Cross-linguistic influence in language learning.* Cambridge: Cambridge University Press.

Ojima, S., Nakata, H., & Kakigi, R. (2005). An ERP study of second language learning after childhood: Effects of proficiency. *Journal of Cognitive Neuroscience*, *17*(8), 1212–28.

Ortega, L. (2009). *Understanding Second Language Acquisition* (p. 304). London: Hodder Education.

Osterhout, L., Allen, M., & McLaughlin, J. (2002). Words in the brain: Lexical determinants of word-induced brain activity. *Journal of Neurolinguistics*, *15*, 171–187.

Osterhout, L., McLaughlin, J., Pitkänen, I., Frenck-Mestre, C., & Molinaro, N. (2006). Novice learners, longitudinal designs, and event-related potentials: A means for exploring the neurocognition of second language processing. *Language Learning*, *56*(1), 199–230.

Otten, M., & Van Berkum, J. J. A. (2009). Does working memory capacity affect the ability to predict upcoming words in discourse? *Brain Research*, *1291*, 92–101.

Oyama, S. (1976). A sensitive period for the acquisition of a nonnative phonological system. *Journal of Psycholinguistic Research*, *5*(3), 261–283.

Pakulak, E., & Neville, H. J. (2010). Proficiency differences in syntactic processing of monolingual native speakers indexed by event-related potentials. *Journal of Cognitive Neuroscience*, *22*(12), 2728–44.

Pakulak, E., & Neville, H. J. (2011). Maturational constraints on the recruitment of early processes for syntactic processing. *Journal of Cognitive Neuroscience*, *23*(10), 2752–2765.

Paolieri, D., Cubelli, R., Macizo, P., Bajo, T., Lotto, L., & Job, R. (2010). Grammatical gender processing in Italian and Spanish bilinguals. *The Quarterly Journal of Experimental Psychology*, *63*(8), 1–15.

Paolieri, D., Lotto, L., Leoncini, D., Cubelli, R., & Job, R. (2011). Differential effects of grammatical gender and gender inflection in bare noun production. *British Journal of Psychology*, *102*, 19–36.

Paolieri, D., Lotto, L., Morales, L., Bajo, T., Cubelli, R., & Job, R. (2010). Grammatical gender processing in romance languages: Evidence from bare noun production in Italian and Spanish. *European Journal of Cognitive Psychology*, *22*(3), 335–347.

Papadopoulou, D., & Clahsen, H. (2003). Parsing strategies in L1 and L2 sentence processing: A study of relative clause attachment in Greek. *Studies in Second Language Acquisition*, *25*, 501–528.

Paradis, M. (2004). *A Neurolinguistic Theory of Bilingualism* (Vol. 18). Amsterdam: John Benjamins.

Patkowski, M. S. (1980). The sensitive period for the acquisition of syntax in a second language. *Language Learning*, *30*(2), 449 – 472.

Penke, M., & Krause, M. (2002). German noun plurals: A challenge to the Dual-Mechanism Model. *Brain and Language*, *81*(1-3), 303–311.

Perani, D. (2005). The neural basis of language talent in bilinguals. *Trends in Cognitive Sciences*, *9*(5), 211–3.

Perani, D., & Abutalebi, J. (2005). The neural basis of first and second language processing. *Current Opinion in Neurobiology*, *15*, 202–206.

Perani, D., Dehaene, S., Grassi, F., Cohen, L., Cappa, S. F., Dupoux, E., … Mehler, J. (1996). Brain processing of native and foreign langauges. *NeuroReport*, *7*, 2439–2444.

Perani, D., Paulesu, E., Galles, N. S., Dupoux, E., Dehaene, S., Bettinardi, V., … Mehler, J. (1998). The bilingual brain. Proficiency and age of acquisition of the second language. *Brain*, *121*, 1841–52.

Phillips, C., Kazanina, N., & Abada, S. H. (2005). ERP effects of the processing of syntactic long-distance dependencies. *Cognitive Brain Research*, *22*(3), 407–28.

Pienemann, M., Di Biase, B., Kawaguchi, S., & Håkansson, G. (2005). Processing constraints on L1 transfer. In Judith F. Kroll & A. M. B. de Groot (Eds.), *Handbook of Bilingualism: Psycholinguistic Approaches* (pp. 128–153). Oxford: Oxford University Press.

Pinker, S., & Ullman, M. T. (2002). The past-tense debate: The past and future of the past tense. *Trends in Cognitive Sciences*, *6*(11), 456–463.

Piske, T., Mackay, I. R. A., & Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics*, *29*, 191–215.

Prado, E. L., & Ullman, M. T. (2009). Can imageability help us draw the line between storage and composition? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(4), 849–66.

Pylkkänen, L., Oliveri, B., & Smart, A. J. (2009). Semantics vs. world knowledge in prefrontal cortex. *Language and Cognitive Processes*, *24*(9), 1313–1334.

Quasthoff, U., Richter, M., & Biemann, C. (2006). Corpus portal for search in monolingual corpora. In *Proceedings of the LREC 2006, Genova, Italy* (pp. 1799–1802).

Reumuth, W., & Winkelmann, O. (1997). *Praktische Grammatik der spanischen Sprache* (p. 356). Wilhelmsfeld: Gottfried Egert Verlag.

Rogers, M. (1987). Learners' difficulties with grammatical gender in German as a foreign language. *Applied Linguistics*, *8*(1), 48–74.

Rossi, S., Gugler, M. F., Friederici, A. D., & Hahne, A. (2006). The impact of proficiency on syntactic second-language processing of German and Italian: Evidence from event-related potentials. *Journal of Cognitive Neuroscience*, *18*(12), 2030–48.

Rugg, M. D. (1990). Event-related brain potentials dissociate repetition effects of high- and low-frequency words. *Memory & Cognition*, *18*(4), 367–79.

Sabourin, L., & Stowe, L. (2004). Memory effects in syntactic ERP tasks. *Brain and Cognition*, *55*(2), 392–5.

Sabourin, L., & Stowe, L. A. (2008). Second language processing: When are first and second languages processed similarly? *Second Language Research*, *24*(3), 397–430.

Sabourin, L., Stowe, L. A., & de Haan, G. J. (2006). Transfer effects in learning a second language grammatical gender system. *Second Language Research*, *22*(1), 1–29.

Sagarra, N., & Herschensohn, J. (2010). The role of proficiency and working memory in gender and number agreement processing in L1 and L2 Spanish. *Lingua*, *120*(8), 2022–2039.

Sagarra, N., & Herschensohn, J. (2011a). Proficiency and animacy effects on L2 gender agreement processes during comprehension. *Language Learning*, *61*(1), 80–116.

Sagarra, N., & Herschensohn, J. (2011b). Asymmetries in gender and number agreement processing in late bilinguals. In *13th Hispanic Linguistics Symposium* (pp. 169–177).

Salamoura, A., & Williams, J. N. (2007a). Processing verb argument structure across languages: Evidence for shared representations in the bilingual lexicon. *Applied Psycholinguistics*, *28*(04), 627–660.

Salamoura, A., & Williams, J. N. (2007b). The representation of grammatical gender in the bilingual lexicon: Evidence from Greek and German. *Bilingualism: Language and Cognition*, *10*(03), 257–275.

Sanders, L. D., & Neville, H. J. (2003). An ERP study of continuous speech processing. II. Segmentation, semantics, and syntax in non-native speakers. *Cognitive Brain Research*, *15*(3), 214–27.

Sasaki, Y. (1991). English and Japanese interlanguage comprehension strategies: An analysis based on the competition model. *Applied Psycholinguistics*, *12*, 47.

Scherag, A., Demuth, L., Rösler, F., Neville, H. J., & Röder, B. (2004). The effects of late acquisition of L2 and the consequences of immigration on L1 for semantic and morpho-syntactic language aspects. *Cognition*, *93*(3), B97–108.

Scheutz, M. J., & Eberhard, K. M. (2004). Effects of morphosyntactic gender features in bilingual language processing. *Cognitive Science*, *28*(4), 559–588.

Schiller, N. O., & Caramazza, A. (2003). Memory and language grammatical feature selection in noun phrase production: Evidence from German and Dutch. *Journal of Memory and Language*, *48*, 169–194.

Schmitt, B. M., Lamers, M., & Münte, T. F. (2002). Electrophysiological estimates of biological and syntactic gender violation during pronoun processing. *Cognitive Brain Research*, *14*(3), 333–46.

Schriefers, H. J. (1993). Syntactic processes in the production of noun phrases. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(4), 841–850.

Schriefers, H. J., & Jescheniak, J. D. (1999). Representation and processing of grammatical gender in language production: A review. *Journal of Psycholinguistic Research*, *28*(6), 575–600.

Schriefers, H. J., & Teruel, E. (2000). Grammatical gender in noun phrase production: The gender interference effect in German. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(6), 1368–77.

Schwartz, B. D., & Sprouse, R. A. (1996). L2 cognitive states and the full transfer/full access model. *Second Language Research*, *12*(1), 40–72.

Scovel, T. (1969). Foreign accents, language acquisition, and cerebral dominance. *Language Learning*, *19*(245-253).

Sebastián-Gallés, N., Martí, M. A., Cuetos, F., & Carreiras, M. (2000). *LEXESP: Base de datos informatizada de la lengua española.* Barcelona: Edicions de la Universitat de Barcelona.

Sera, M. D., Elieff, C., Forbes, J., Burch, M. C., Rodríguez, W., & Dubois, D. P. (2002). When language affects cognition and when it does not: An analysis of grammatical gender and classification. *Journal of Experimental Psychology: General*, *131*(3), 377–397.

Sereno, S. C., Brewer, C. C., & O'Donnell, P. J. (2003). Context effects in word recognition: Evidence for early interactive processing. *Psychological Science*, *14*(4), 328–333.

Shin, J.-A., & Christianson, K. (2009). Syntactic processing in Korean-English bilingual production: Evidence from cross-linguistic structural priming. *Cognition*, *112*(1), 175–80.

Slabakova, R. (2000). L1 transfer revisited: The L2 acquisition of telicity marking in English by Spanish and Bulgarian native speakers. *Linguistics*, *38*(4), 739–770.

Sorace, A. (1993). Incomplete vs. divergent representations of unaccusativity in non native grammars of Italian. *Second Language Research*, *9*(1), 22–47.

Spalding, D. A. (1872). Instinct. With original observations on young animals. *Macmillan´s Magazine*, *27*, 282–293.

Spinner, P., & Juffs, A. (2008). L2 grammatical gender in a complex morphological system: The case of German. *IRAL - International Review of Applied Linguistics in Language Teaching*, *46*(2008), 315–348.

Steinhauer, K., White, E. J., & Drury, J. E. (2009). Temporal dynamics of late second language acquisition: Evidence from event-related brain potentials. *Second Language Research*, *25*(1), 13–41.

Su, I.-R. (2001). Transfer of sentence processing strategies: A comparison of L2 learners of Chinese and English. *Applied Psycholinguistics*, *22*(1), 83–112.

Szekely, A., Devescovi, A., Federmeier, K. D., Herron, D., Iyer, G., Jacobsen, T., … Bates, E. A. (2005). Timed action and object naming. *Cortex*, *41*(1), 7–25.

Szekely, A., Jacobsen, T., D'Amico, S., Devescovi, A., Andonova, E., Herron, D., … Bates, E. A. (2004). A new on-line resource for psycholinguistic studies. *Journal of Memory and Language*, *51*(2), 247–250.

Taraban, R., McDonald, J. L., & MacWhinney, B. (1989). Category learning in a connectionist model: Learning to decline the German definite article. In R. Corrigan, F. Eckman, & M. Noonan (Eds.), *Linguistic Categorization* (pp. 163–193). Amsterdam: John Benjamins.

Tokowicz, N., & MacWhinney, B. (2005). Implicit and explicit measures of sensitivity to violations in second language grammar: An Event-Related Potential Investigation. *Studies in Second Language Acquisition*, *27*, 173–204.

Ullman, M. T. (1999). Acceptability ratings of regular and irregular past-tense forms: Evidence for a dual-system model of language from word frequency and phonological neighbourhood effects. *Language and Cognitive Processes*, *14*(1), 47–67.

Ullman, M. T. (2001a). The declarative/procedural model of lexicon and grammar. *Journal of Psycholinguistic Research*, *30*(1), 37–69.

Ullman, M. T. (2001b). The neural basis of lexicon and grammar in first and second language: The declarative/procedural model. *Bilingualism: Language and Cognition*, *4*(1), 105–122.

Ullman, M. T. (2001c). A neurocognitive perspective on language: The declarative/procedural model. *Nature Reviews Neuroscience*, *2*, 717–26.

Ullman, M. T. (2004). Contributions of memory circuits to language: The declarative/procedural model. *Cognition*, *92*(1-2), 231–70.

Ullman, M. T. (2012). Declarative/Procedural Model (DP). In P. Robinson (Ed.), *Routledge Encyclopedia of Second Language Acquisition* (pp. 160–164). New York: Routledge.

Ullman, M. T., Corkin, S., Hickok, G., Growdon, J. H., Koroshetz, W. J., & Pinker, S. (1997). A neural dissociation within language: Evidence that the mental dictionary is part of declarative memory, and that grammatical rules are processed by the procedural system. *Journal of Cognitive Neuroscience*, *9*(2), 266–276.

Ullman, M. T., & Gopnik, M. (1999). Inflectional morphology in a family with inherited specific language impairment. *Applied Psycholinguistics*, *20*(01), 51–117.

Ullman, M. T., Pancheva, R., Love, T., Yee, E., Swinney, D., & Hickok, G. (2005). Neural correlates of lexicon and grammar: evidence from the production, reading, and judgment of inflection in aphasia. *Brain and Language*, *93*(2), 185–238; discussion 239–42.

Ullman, M. T., & Walenski, M. (2005). Moving past the past tense. *Brain and Language*, *93*(2), 248–252.

Van Casteren, M., & Davis, M. H. (2006). Mix, a program for pseudorandomization. *Behavior Research Methods*, *38*(4), 584–9.

Van Casteren, M., & Davis, M. H. (2007). Match: A program to assist in matching the conditions of factorial experiments. *Behavior Research Methods*, *39*(4), 973–8.

Van der Lely, H. K. J., & Ullman, M. T. (2001). Past tense morphology in specifically language impaired and normally developing children. *Language and Cognitive Processes*, *16*(2-3), 177–217.

Van Hell, J. G., & Dijkstra, T. (2002). Foreign language knowledge can influence native language performance in exclusively native contexts. *Psychonomic Bulletin & Review*, *9*(4), 780–789.

Van Heuven, W. J. B., Dijkstra, T., & Grainger, J. (1998). Orthographic neighborhood effects in bilingual word recognition. *Cognition*, *483*, 458–483.

Van Petten, C., & Kutas, M. (1990). Interactions between sentence context and word frequency in event-related brain potentials. *Memory & Cognition*, *18*(4), 380–93.

Vera Morales, J. (2013). *Spanische Grammatik* (p. 826). München: Oldenbourg Verlag.

Wartenburger, I., Heekeren, H. R., Abutalebi, J., Cappa, S. F., Villringer, A., & Perani, D. (2003). Early setting of grammatical processing in the bilingual brain. *Neuron*, *37*(1), 159–70.

Weber, A., & Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language*, *50*, 1–25.

Weber, K., & Indefrey, P. (2009). Syntactic priming in German-English bilinguals during sentence comprehension. *NeuroImage*, *46*(4), 1164–72.

Weber, K., & Lavric, A. (2008). Syntactic anomaly elicits a lexico-semantic (N400) ERP effect in the second language but not the first. *Psychophysiology*, *45*(6), 920–5.

Weber-Fox, C. M., Davis, L. J., & Cuadrado, E. (2003). Event-related brain potential markers of high-language proficiency in adults. *Brain and Language*, *85*(2), 231–244.

Weber-Fox, C. M., & Neville, H. J. (1996). Maturational constraints onfunctional specializations for language processing: ERP and behavioral evidence in bilingual speakers. *Journal of Cognitive Neuroscience*, *8*(3), 231 – 256.

Weber-Fox, C., & Neville, H. J. (2001). Sensitive periods differentiate processing of open- and closed-class words. An ERP study of bilinguals. *Journal of Speech, Language, and Hearing Research*, *44*, 1338–1535.

White, L. (1989). *Universal Grammar and Second Language Acquisition* (p. 198). Amsterdam: John Benjamins.

White, L., & Genesee, F. (1996). How native is near-native? The issue of ultimate attainment in adult second language acquisition. *Second Language Research*, *12*(3), 233–265.

White, L., Valenzuela, E., Kozlowska–Macgregor, M., & Leung, Y.-K. I. (2004). Gender and number agreement in nonnative Spanish. *Applied Psycholinguistics*, *25*(01), 105–133.

Wicha, N. Y. Y., Moreno, E. M., & Kutas, M. (2004). Anticipating words and their gender: An event-related brain potential study of semantic integration, gender expectancy, and gender agreement in Spanish sentence reading. *Journal of Cognitive Neuroscience*, *16*(7), 1272–1288.

Wiley, E. W., Bialystok, E., & Hakuta, K. (2005). New approaches to using census data to test the critical-period hypothesis for second-language acquisition. *Psychological Science*, *16*(4), 341–3.

Williams, J. N., Möbius, P., & Kim, C. (2001). Native and non-native processing of English wh- questions: Parsing strategies and plausibility constraints. *Applied Psycholinguistics*, *22*(04), 1–28.

8. List of tables

9. List of figures

10. Appendices

10.1 Material Experiment 1

Noncognates

Congruent condition

	German article	German noun	Spanish article	Spanish noun	English translation
1.	die	Spinne	la	araña	spider
2.	die	Fahne	la	bandera	flag
3.	die	Flasche	la	botella	bottle
4.	die	Stadt	la	ciudad	town
5.	die	Milch	la	leche	milk
6.	die	Wurst	la	salchicha	sausage
7.	die	Schlange	la	serpiente	snake
8.	die	Puppe	la	muñeca	doll
9.	die	Kette	la	cadena	chain
10.	die	Ziege	la	cabra	goat
11.	die	Kuh	la	vaca	cow
12.	die	Träne	la	lágrima	tear
13.	der	Hammer	el	martillo	hammer
14.	der	Schuh	el	zapato	shoe
15.	der	Vogel	el	pájaro	bird
16.	der	Knochen	el	hueso	bone
17.	der	Hund	el	perro	dog
18.	der	Kreis	el	círculo	circle
19.	der	Fisch	el	pez	fish
20.	der	Nabel	el	ombligo	navel
21.	der	Boden	el	suelo	floor
22.	der	Knoten	el	nudo	knot
23.	der	Bär	el	oso	bear
24.	der	Ring	el	anillo	ring

Incongruent condition

	German article	German noun	Spanish article	Spanish noun	English translation
1.	der	Löffel	la	cuchara	spoon
2.	der	Pfeil	la	flecha	arrow
3.	der	Apfel	la	manzana	apple
4.	der	Tropfen	la	gota	drop
5.	der	Mund	la	boca	mouth
6.	der	Stuhl	la	silla	chair
7.	der	Stein	la	piedra	stone
8.	der	Besen	la	escoba	broom
9.	der	Regen	la	lluvia	rain
10.	der	Koffer	la	maleta	suitcase
11.	der	Ast	la	rama	branch
12.	der	Berg	la	montaña	mountain
13.	die	Bürste	el	cepillo	brush
14.	die	Brust	el	pecho	chest
15.	die	Wüste	el	desierto	desert
16.	die	Schulter	el	hombro	shoulder
17.	die	Welt	el	mundo	world
18.	die	Maus	el	ratón	mouse
19.	die	Uhr	el	reloj	watch
20.	die	Ampel	el	semáforo	stop light
21.	die	Schraube	el	tornillo	screw
22.	die	Vase	el	florero	vase
23.	die	CD	el	disco	CD
24.	die	Hose	el	pantalón	pants

Incongruent neutral condition

	German article	German noun	Spanish article	Spanish noun	English translation
1.	das	Fahrrad	la	bicicleta	bike
2.	das	Bett	la	cama	bed
3.	das	Schaf	la	oveja	sheep
4.	das	Blatt	la	hoja	leaf
5.	das	Fleisch	la	carne	meat
6.	das	Bier	la	cerveza	beer
7.	das	Kissen	la	almohada	pillow
8.	das	Hemd	la	camisa	shirt
9.	das	Seil	la	cuerda	rope
10.	das	Bein	la	pierna	leg
11.	das	Knie	la	rodilla	knee
12.	das	Gemüse	la	verdura	vegetable
13.	das	Gehirn	el	cerebro	brain
14.	das	Messer	el	cuchillo	knife
15.	das	Klavier	el	piano	piano
16.	das	Kleid	el	vestido	dress
17.	das	Buch	el	libro	book
18.	das	Brot	el	pan	bread
19.	das	Pferd	el	caballo	horse
20.	das	Dreieck	el	triángulo	triangle
21.	das	Heft	el	cuaderno	notebook
22.	das	Gewehr	el	fusil	rifle
23.	das	Geschenk	el	regalo	gift
24.	das	Eis	el	helado	ice cream

Cognates

Congruent condition

	German article	German noun	Spanish article	Spanish noun	English translation
1.	die	Tasse	la	taza	cup
2.	die	Gitarre	la	guitarra	guitar
3.	die	Palme	la	palmera	palm tree
4.	der	Mais	el	maíz	corn
5.	der	Tunnel	el	túnel	tunnel
6.	der	Tiger	el	tigre	tiger
7.	die	Pyramide	la	pirámide	pyramid
8.	die	Trompete	la	trompeta	trumpet
9.	der	Motor	el	motor	motor
10.	die	Pistole	la	pistola	pistol
11.	die	Kamera	la	cámara	camera
12.	der	Elefant	el	elefante	elephant
13.	der	Satellit	el	satélite	satellite
14.	der	Roboter	el	robot	robot

Incongruent condition

	German article	German noun	Spanish article	Spanish noun	English translation
1.	der	Salat	la	ensalada	salad
2.	der	Computer	la	computadora	computer
3.	der	Turm	la	torre	tower
4.	die	Katze	el	gato	cat
5.	die	Garage	el	garaje	garage
6.	die	Melone	el	melón	melon
7.	die	Schokolade	el	chocolate	chocolate
8.	der	Bart	la	barba	beard
9.	der	Hamburger	la	hamburguesa	hamburger
10.	die	Zigarette	el	cigarro	cigarette
11.	die	Sonne	el	sol	sun
12.	die	Tomate	el	tomate	tomato
13.	die	Uniform	el	uniforme	uniform
14.	die	Avocado	el	aguacate	avocado

Incongruent neutral condition

	German article	German noun	Spanish article	Spanish noun	English translation
1.	das	Atom	el	átomo	atom
2.	das	Mikrofon	el	micrófono	microphone
3.	das	Salz	la	sal	salt
4.	das	Kabel	el	cable	cable
5.	das	Kamel	el	camello	camel
6.	das	Saxophon	el	saxofón	saxophone
7.	das	Hotel	el	hotel	hotel
8.	das	Mikroskop	el	microscopio	microscope
9.	das	Baby	el	bebe	baby
10.	das	Stadion	el	estadio	stadium
11.	das	Orchester	la	orquesta	orchestra
12.	das	Parfüm	el	perfume	perfume
13.	das	Paket	el	paquete	packet
14.	das	Taxi	el	taxi	taxi

10.2 Material Experiment 2

Experimental sentences

	Introductory sentence			Second sentence				
1.	This	is	a crown.	*He/*She/It	is	shiny	and	golden.
2.	This	is	a uniform.	*He/*She/It	is	black	and	elegant.
3.	This	is	a blouse.	*He/*She/It	is	modern	and	feminine.
4.	This	is	a jacket.	*He/*She/It	is	warm	and	heavy.
5.	This	is	a sock.	*He/*She/It	is	simple	and	woollen.
6.	This	is	a mattress.	*He/*She/It	is	hard	and	uncomfortable.
7.	This	is	an orange.	*He/*She/It	is	juicy	and	tasty.
8.	This	is	a banana.	*He/*She/It	is	yellow	and	sweet.
9.	This	is	a tomato.	*He/*She/It	is	ripe	and	soft.
10.	This	is	a plant.	*He/*She/It	is	green	and	pretty.
11.	This	is	a rose.	*He/*She/It	is	rare	and	beautiful.
12.	This	is	a bakery.	*He/*She/It	is	clean	and	tidy.
13.	This	is	a guitar.	*He/*She/It	is	antique	and	valuable.
14.	This	is	a trumpet.	*He/*She/It	is	small	and	loud.
15.	This	is	a cigarette.	*He/*She/It	is	white	and	orange.
16.	This	is	a feather.	*He/*She/It	is	grey	and	purple.
17.	This	is	a lamp.	*He/*She/It	is	new	and	cheap.
18.	This	is	a line.	*He/*She/It	is	short	and	thick.
19.	This	is	a machine.	*He/*She/It	is	complicated	and	unsafe.
20.	This	is	a street.	*He/*She/It	is	narrow	and	dangerous.
21.	This	is	a rocket.	*He/*She/It	is	fast	and	powerful.
22.	This	is	a school.	*He/*She/It	is	good	and	interesting.
23.	This	is	a university.	*He/*She/It	is	progressive	and	different.
24.	This	is	an island.	*He/*She/It	is	empty	and	boring.
25.	This	is	a finger.	*He/*She/It	is	long	and	skinny.
26.	This	is	an arm.	*He/*She/It	is	strong	and	brown.
27.	This	is	a foot.	*He/*She/It	is	big	and	hairy.
28.	This	is	a ring.	*He/*She/It	is	silver	and	fragile.
29.	This	is	a shoe.	*He/*She/It	is	wet	and	dirty.
30.	This	is	a helmet.	*He/*She/It	is	light	and	safe.
31.	This	is	a hat.	*He/*She/It	is	colorful	and	crazy.
32.	This	is	a seat.	*He/*She/It	is	old	and	comfortable.
33.	This	is	a salad.	*He/*She/It	is	fresh	and	delicious.
34.	This	is	an apple.	*He/*She/It	is	red	and	healthy.
35.	This	is	a balcony.	*He/*She/It	is	sunny	and	hot.
36.	This	is	a park.	*He/*She/It	is	shady	and	pleasant.
37.	This	is	a garden.	*He/*She/It	is	quiet	and	relaxing.
38.	This	is	a bush.	*He/*She/It	is	dry	and	ugly.
39.	This	is	a market.	*He/*She/It	is	famous	and	popular.
40.	This	is	a circus.	*He/*She/It	is	funny	and	entertaining.
41.	This	is	a hammer.	*He/*She/It	is	professional	and	expensive.
42.	This	is	a ball.	*He/*She/It	is	blue	and	pink.
43.	This	is	a planet.	*He/*She/It	is	unknown	and	strange.
44.	This	is	a bus.	*He/*She/It	is	slow	and	crowded.
45.	This	is	a stone.	*He/*She/It	is	round	and	smooth.
46.	This	is	a canal.	*He/*She/It	is	deep	and	wide.
47.	This	is	a tunnel.	*He/*She/It	is	cold	and	dark.
48.	This	is	a fish.	*He/*She/It	is	thin	and	quick.

Filler sentences

a) Grammatical violations fillers

	Introductory sentence			Second sentence				
1.	This	is	a bar.	*It	are	smoky	and	noisy.
2.	This	is	a bank.	*It	are	huge	and	spacious.
3.	This	is	a computer.	*It	are	dusty	and	broken.
4.	This	is	a beard.	*It	are	blond	and	curly.
5.	This	is	a photo.	*It	are	artistic	and	bizarre.
6.	This	is	a beer.	*It	are	cool	and	refreshing.
7.	This	is	a hotel.	*It	are	unique	and	special.
8.	This	is	a book.	*It	are	difficult	and	bad.

b) People fillers

	Introductory sentence			Second sentence				
1.	This	is	a beautician.	He/She/*It	is	hungry	and	thirsty.
2.	This	is	an au pair.	He/She/*It	is	tired	and	forgetful.
3.	This	is	a secretary.	He/She/*It	is	angry	and	sad.
4.	This	is	a hairdresser.	He/She/*It	is	calm	and	relaxed.
5.	This	is	a dressmaker.	He/She/*It	is	confident	and	intelligent.
6.	This	is	a nurse.	He/She/*It	is	happy	and	energetic.
7.	This	is	an ice skater.	He/She/*It	is	fair	and	generous.
8.	This	is	a flight attendant.	He/She/*It	is	polite	and	friendly.
9.	This	is	a model.	He/She/*It	is	young	and	athletic.
10.	This	is	a dancer.	He/She/*It	is	rude	and	direct.
11.	This	is	a sales assistant.	He/She/*It	is	married	and	jealous.
12.	This	is	a librarian.	He/She/*It	is	patient	and	careful.
13.	This	is	a politician.	He/She/*It	is	talented	and	clever.
14.	This	is	a lorry driver.	He/She/*It	is	excited	and	nervous.
15.	This	is	a chemist.	He/She/*It	is	lonely	and	impolite.
16.	This	is	a judge.	He/She/*It	is	nice	and	understanding.
17.	This	is	a butcher.	He/She/*It	is	worried	and	unhappy.
18.	This	is	a mechanic.	He/She/*It	is	sick	and	poor.
19.	This	is	an electrician.	He/She/*It	is	religious	and	proud.
20.	This	is	a farmer.	He/She/*It	is	weak	and	lazy.
21.	This	is	an engineer.	He/She/*It	is	pale	and	afraid.
22.	This	is	a technician.	He/She/*It	is	qualified	and	responsible.
23.	This	is	a boxer.	He/She/*It	is	conservative	and	serious.
24.	This	is	a pilot.	He/She/*It	is	cheerful	and	active.

c) Correct fillers

	Introductory sentence			Second sentence				
1.	There	are	two cats.	They	are	purring	on	the sofa.
2.	There	are	three bees	They	are	humming	in	the sun.
3.	There	are	four spiders.	They	are	weaving	a large	web.
4.	There	are	five ants.	They	are	carrying	a piece of	wood.
5.	There	are	two bells.	They	are	ringing	in	the church.
6.	There	are	three blankets.	They	are	lying	on	the chair.
7.	There	are	four bottles.	They	are	falling	from	the shelf.
8.	There	are	five candles.	They	are	glowing	on	the table.
9.	There	are	two dolls.	They	are	sitting	on	the bench.
10.	There	are	three towns.	They	are	buzzing	with	activity.
11.	There	are	four clouds	They	are	moving	in	the sky.
12.	There	are	five ducks.	They	are	swimming	in	the lake.
13.	There	are	two snakes.	They	are	crawling	over	the ground.
14.	There	are	three goats.	They	are	grazing	on	the meadow.
15.	There	are	four pears.	They	are	rotting	in	the cellar.
16.	There	are	five flowers.	They	are	blooming	in	the yard.
17.	There	are	two dragons.	They	are	fighting	against	the knight.
18.	There	are	three dinosaurs.	They	are	attacking	another	dinosaur.
19.	There	are	four lions.	They	are	chasing	after	an antelope.
20.	There	are	five dolphins.	They	are	diving	in	the sea.
21.	There	are	two helicopters.	They	are	crashing	into	a building.
22.	There	are	three parrots.	They	are	nesting	in	a tree.
23.	There	are	four penguins.	They	are	sliding	down	the iceberg.
24.	There	are	five wolves.	They	are	hunting	in	the woods.
25.	There	are	two dogs.	They	are	barking	at	the thief.
26.	There	are	three donkeys.	They	are	returning	to	the stable.
27.	There	are	four birds.	They	are	flying	over	a mountain.
28.	There	are	five stars.	They	are	twinkling	in	the night.
29.	There	are	two arrows.	They	are	pointing	to	the left.
30.	There	are	three trains.	They	are	arriving	at	the station.
31.	There	are	four closets.	They	are	standing	in	the hallway.
32.	There	are	five monkeys.	They	are	climbing	in	their cage.
33.	There	are	two canoes.	They	are	floating	on	the river.
34.	There	are	three crocodiles.	They	are	yawning	with	boredom.
35.	There	are	four kangaroos.	They	are	jumping	in	Australia.
36.	There	are	five camels.	They	are	crossing	through	the desert.
37.	There	are	two monsters.	They	are	hiding	under	the bed.
38.	There	are	three sheep.	They	are	sleeping	in	the barn.
39.	There	are	four lambs.	They	are	playing	in	the grass.
40.	There	are	five zebras.	They	are	looking	for	water.
41.	There	are	two flags.	They	are	waving	in	the wind.
42.	There	are	three taxis.	They	are	waiting	for	customers.
43.	There	are	four chicken.	They	are	pecking	in	the garbage.
44.	There	are	five horses.	They	are	running	in	the prairie.
45.	There	are	two cars.	They	are	racing	on	the highway.
46.	There	are	three pigs.	They	are	rolling	in	the mud.
47.	There	are	four towels.	They	are	dripping	on	the floor.
48.	There	are	five rabbits.	They	are	hopping	through	the fields.